# JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 8 (2020), Issue 5 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

## Contents

## Corrigenda and Addendas

Viewpoint

# Toward Earlier Diagnosis Using Combined eHealth Tools in Rheumatology: The Joint Pain Assessment Scoring Tool (JPAST) Project

Johannes Knitza[1], MD; Rachel Knevel[2,3], MD; Karim Raza[4,5], MD; Tor Bruce[6], BAPSC; Ekaterina Eimer[7], Dr rer nat; Isabel Gehring[7], Dr rer nat; Linda Mathsson-Alm[8], PhD; Maryam Poorafshar[8], PhD; Axel J Hueber[1], MD; Georg Schett[1], MD; Martina Johannesson[9], PhD; Anca Catrina[9], MD; Lars Klareskog[9], MD; JPAST Group[9]

[1]Department of Internal Medicine 3 – Rheumatology and Immunology, Friedrich-Alexander University Erlangen-Nürnberg, University Hospital Erlangen, Erlangen, Germany

[2]Department of Rheumatology, Leiden University Medical Center, Leiden, Netherlands

[3]Division of Rheumatology, Inflammation and Immunity, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

[4]Department of Rheumatology, Sandwell and West, Birmingham Hospitals NHS Trust, Birmingham, United Kingdom

[5]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, United Kingdom

[6]Ocean Observations, Stockholm, Sweden

[7]Thermo Fisher Scientific, Freiburg, Germany

[8]Thermo Fisher Scientific, Uppsala, Sweden

[9]Division of Rheumatology, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

**Corresponding Author:**
Johannes Knitza, MD
Department of Internal Medicine 3 – Rheumatology and Immunology
Friedrich-Alexander University Erlangen-Nürnberg
University Hospital Erlangen
Ulmenweg 18
Erlangen, 91054
Germany
Phone: 49 91318532093
Fax: 49 91318532092
Email: johannes.knitza@uk-erlangen.de

## Abstract

Outcomes of patients with inflammatory rheumatic diseases have significantly improved over the last three decades, mainly due to therapeutic innovations, more timely treatment, and a recognition of the need to monitor response to treatment and to titrate treatments accordingly. Diagnostic delay remains a major challenge for all stakeholders. The combination of electronic health (eHealth) and serologic and genetic markers holds great promise to improve the current management of patients with inflammatory rheumatic diseases by speeding up access to appropriate care. The Joint Pain Assessment Scoring Tool (JPAST) project, funded by the European Union (EU) European Institute of Innovation and Technology (EIT) Health program, is a unique European project aiming to enable and accelerate personalized precision medicine for early treatment in rheumatology, ultimately also enabling prevention. The aim of the project is to facilitate these goals while at the same time, reducing cost for society and patients.

***(JMIR Mhealth Uhealth 2020;8(5):e17507)*** doi:10.2196/17507

**KEYWORDS**

rheumatology; eHealth; mHealth; symptom-checkers; apps

## Background

The path to a correct diagnosis and efficacious treatment is often long and frustrating for patients with inflammatory rheumatic and musculoskeletal diseases (RMDs). This is a major problem as both short and long treatment efficacy depends on early and correct diagnosis. Early diagnosis for rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), myositis, primary Sjögren

syndrome (SS), and systemic sclerosis (SSc) is critical for improved disease outcomes and the selection of a therapy strategy. For example, in the context of RA, the first 3 months of symptoms have been identified as a therapeutic window during which immunologic mechanisms can still be altered [1,2]. Therefore, the European League Against Rheumatism (EULAR) recommends that any patient presenting with morning stiffness or joint pain or swollen joints sees a rheumatologist no later than 6 weeks after symptom onset [3]. Interestingly, the advice to see a rheumatologist for persistent joint pain poses a particular challenge, as joint pain is very common in the population [4]. However, certain variants of joint pain together with the presence of rheumatoid factor (RF) or anticitrullinated protein antibody (ACPA) indicate a high risk for the development of RA. When identified, such individuals can be given lifestyle advice to reduce the risk of disease development and an opportunity to participate in clinical trials aimed at prevention of RA. Notably, novel ways of identifying patients at risk on a large scale will be needed if ongoing preventive trials are successful and will lead to a change in clinical practice.

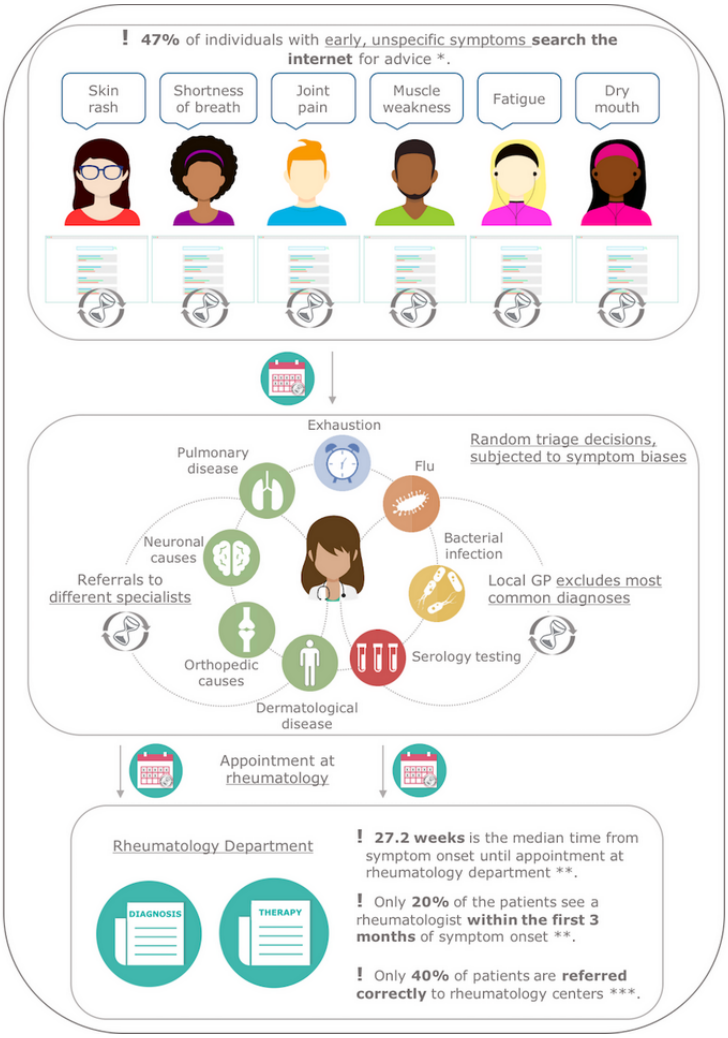
## Diagnostic Delay in Today's Clinical Practice

Diagnostic delay [5] is one of the biggest current challenges in rheumatology. Rheumatic symptoms such as joint pain are common and hard to evaluate for patients and health care providers [6-8]. Patients often wait too long as they believe that the symptoms will resolve spontaneously [9] or with self-care methods [10]. General practitioners (GP) find it hard to identify RMD symptoms indicative of emerging RA or other inflammatory RMD at early stages in the disease course [11]. The resulting delay may exacerbate existing health disparities [10,11]. Furthermore, rheumatologists remain scarce worldwide [12], and this represents one of the main reasons for the delay in diagnosis. Notably, the early and correct identification of patients with emerging rheumatic diseases based solely on clinical evaluation is very challenging even for experienced rheumatologists [13].

## Improving Health Care Efficiency

In addition to the challenge of reducing diagnostic delay, there is an increasing need to optimize health care efficiency. Musculoskeletal complaints account for 21.3% of the years lived with disabilities, with neck pain and low back pain accounting for almost 70% [14]. Up to 60% of patients presenting to rheumatologists in many countries turn out to have no inflammatory rheumatic diseases [15] (Figure 1). Given that the prevalence of musculoskeletal complaints increases with age, the increasing age of most populations will lead to a systemic overload of health care systems. The solution to such situations is to classically triage patients to allow prioritization based on the level of urgency and availability of effective treatment. There are different strategies [16,17] to accelerate access to rheumatologists, although low-barrier electronic health (eHealth)–based approaches remain rare. In emergency departments, heterogeneous triage decisions led to the creation and use of triage standards such as the widely used Manchester-Triage-System [18]. In rheumatology, no triage system has yet been widely accepted [16] and various local systems are being used [19]. The lack of transparent and objective standards for triage [20] represents a major hurdle for early diagnosis in patients where early treatment could make a large difference.

**Figure 1.** Diagnostic delay and inefficient health care service. GP: general practitioner. *Powley et al [21]. **Stack et al [10]. ***Feuchtenberger et al [15].




## The Value of Current Symptom Checkers in Rheumatology

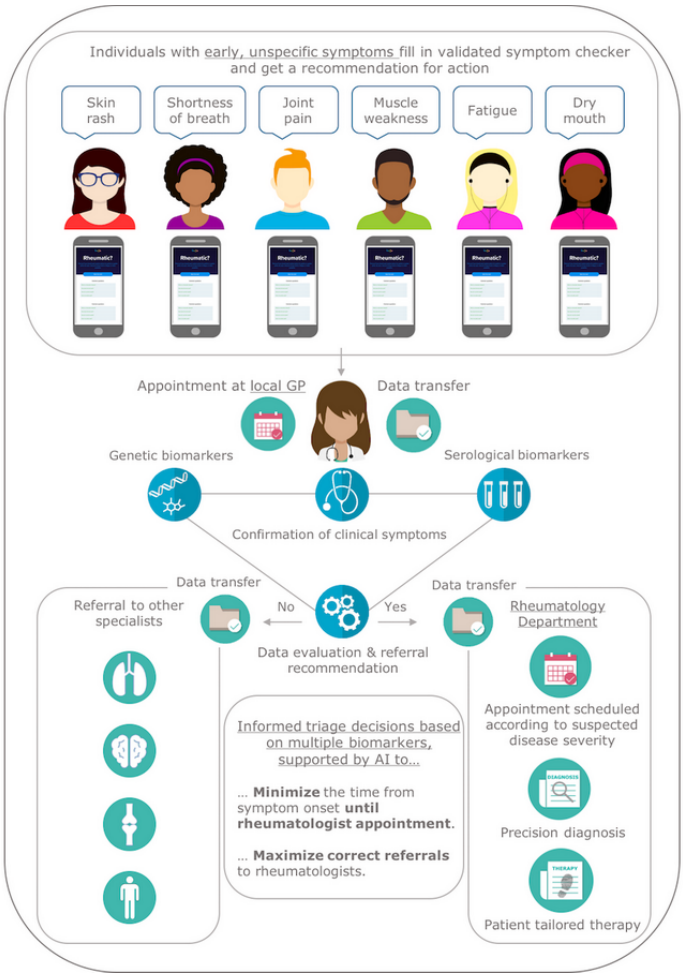
The internet is an important source of information for both health care professionals and members of the public. Patients often check their symptoms online prior to seeing a rheumatologist [21]. Symptom checkers represent a professional alternative to search engines. They represent a patient-facing version of a diagnostic decision support system (DDSS). These systems have existed for a long time, yet are rarely used in clinical practice. Based on the patient's reported symptoms, these systems generate a list of probable diagnoses and offer advice on further steps. These tools can potentially reduce the number of delayed and incorrect diagnoses [22]. Two systematic reviews showed that the diagnostic accuracy of physicians could be improved by using DDSS [23,24]. A retrospective evaluation that applied a DDSS on medical records of patients with rare diseases showed that the median advantage of correct disease suggestions compared to the time of clinical diagnosis was 3 months [25]. The use of symptom checkers in rheumatology has been explored only on a minor level. Powley et al [21] recently evaluated two freely available symptom checkers which included a broad variety of different diseases, and were thus not rheumatology-specific. The study showed that the help-seeking advice and diagnoses given by the symptom checkers tested was frequently inaccurate for rheumatic diseases, correctly identifying inflammatory arthritis in only 4 of 21 (19%) patients. The NHS (National Health Service) symptom checker inappropriately suggested that nearly half of patients should seek advice from emergency services. This phenomenon is well known for symptom checkers, as most systems perform in a relatively risk-adverse manner [26]. Symptom checkers promise to support patients, GPs, and rheumatologists in making the correct diagnosis in a timely manner using minimal resources with little burden; however, more evidence is needed to prove their clinical and economic benefit.

## Incorporation of Serological and Genetic Data

Subjectivity is an inherent weakness of symptom checkers. The possibility to combine the outcome from symptom checkers with serological and genetic biomarkers is likely to drastically improve early differentiation between patients with and without autoimmune inflammatory RMDs. This would shorten the time it takes for patients to get access to specialist care (Figure 2). Several serological biomarkers are included in the American College of Rheumatology (ACR)/EULAR classification criteria

for RMDs (eg, ACPA and RF in RA [27], anti-dsDNA and anti-Sm in SLE [28], Scl70, anti-CENP and anti-RNA polymerase III in SSc [29], anti-Jo-1 in myositis [30] and anti-SSA/Ro in SS [31]). Such biomarkers can be present years before the onset of symptoms [32]. Early diagnostic biomarker panels and biomarkers with predictive utility are needed to guide clinical decision-making in rheumatology. The contribution of genetic variants to the individual risk of patients developing RMDs was highlighted in several twin studies [33,34]. Recently, genome-wide association studies (GWAS) have highlighted specific genes (eg, HLA, STAT4, TNF, PTPN22) for their contribution. Combinations of different genetic risk factors could contribute to the overall differentiation of patients at risk of developing inflammatory RMDs compared with other noninflammatory, nonautoimmune diseases.

**Figure 2.** Shorter patient journey and more efficiency in health care with JPAST (Joint Pain Assessment Scoring Tool). AI: artificial intelligence; GP: general practitioner.




Next-generation technologies have emerged, allowing fast and easy analyses of multiple serology and genetic markers. The early identification of high-risk patients allows them and their physicians to make effective lifestyle and medication changes to reduce the risk of future disease activity. The combination and interrogation of clinical, serological, and genetic data could allow the precise and early separation and description of phenotypes; currently available machine learning approaches offer particular promise in relation to this [35]. The digital documentation of symptoms and complete electronic patient records enable the identification of current treatment bottlenecks, although implementation into the digital infrastructure at the GP level and in different hospital information systems will be a challenge. Using these data, process enhancement will allow the optimization of the current health care situation [36].

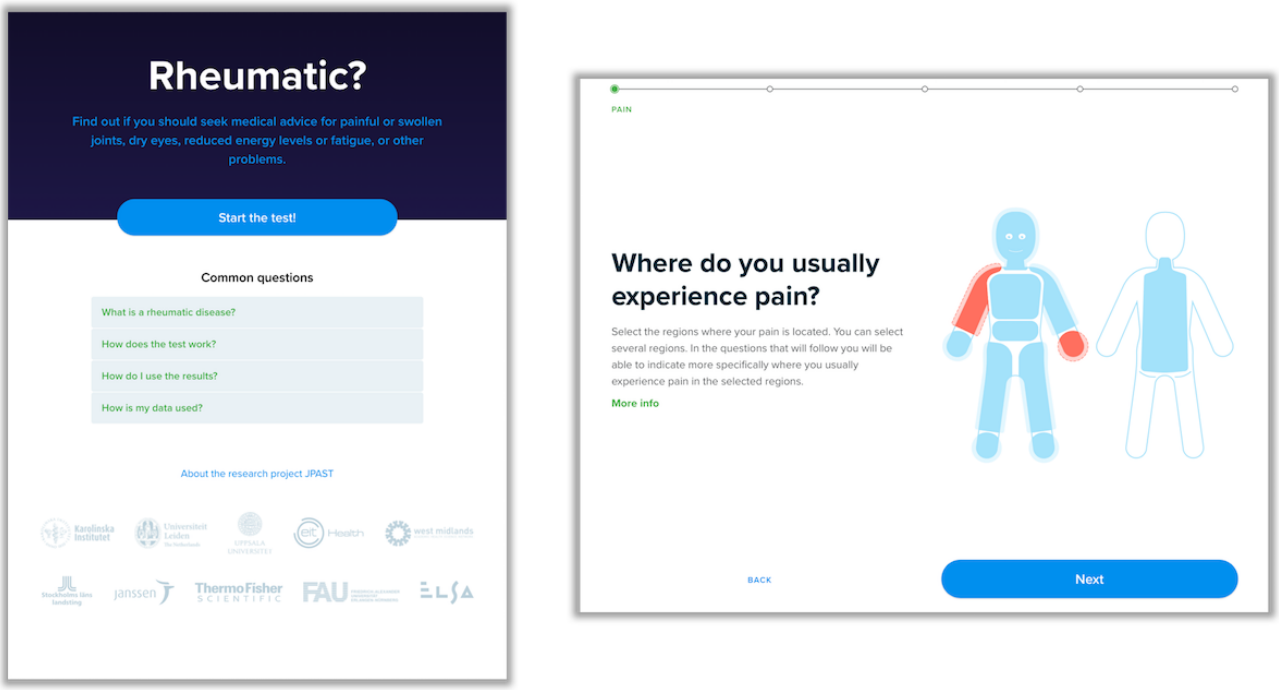
## Joint Pain Assessment Scoring Tool

The objective of the Joint Pain Assessment Scoring Tool (JPAST) project, a digital diagnostic program supported by EU/EIT, is to improve the early differentiation between patients with and without autoimmune inflammatory RMDs and in addition to identify individuals at very high risk for these diseases (Figure 2). To our knowledge JPAST is the first project combining patient symptoms, genetics, and serology biomarkers for these purposes. The intention is that JPAST will eventually be tested at leading university hospitals in Europe and subsequently be implemented elsewhere to accelerate access to rheumatology services and appropriate therapy.

The current version of the JPAST eHealth tool, “Rheumatic?” (Figure 3 and Multimedia Appendix 1), is available in English, Swedish, Dutch, and German. “Rheumatic?” is a website with responsive web design, making it adaptable for smartphone and

tablet use. The interface consists of single-choice buttons, multiple-choice buttons, image area questions, single-choice sliders, and pain sliders (Multimedia Appendix 1). The questions are individualized, as the next question depends on the previous answer. The scoring system was developed to be flexible and adapt to any needs that came up during the project. To achieve this, we created a system of reusable, recursive point buckets. Every point bucket has two threshold values, and a list of assigned points for any number of question options. To evaluate the point bucket, all the selected question options and their total points are calculated. The points can be positive or negative. The sum is compared to the threshold values to see if either threshold should be considered active. The recursive element comes into play when a point bucket assigns points not just directly to one particular option, but also to a threshold in another point bucket. For example, this allows us to create point buckets that are activated on individual symptoms, and then assign different scores to these symptoms for different diseases. This decouples the way points are assigned from the way the questions are asked, and allows more sophisticated scoring logic and more convenient reuse of points. Patients and physicians worked together to develop these tools.

**Figure 3.** Screenshots of "Rheumatic?" the JPAST (Joint Pain Assessment Scoring Tool) symptom checker.




Furthermore, specific serology-based multiplex assays and next generation sequencing panels are also being developed and will be integrated with the “Rheumatic?” eHealth tool. These assays are currently being validated using blood samples from the 5 key RMDs: rheumatoid arthritis, idiopathic inflammatory myopathies, systemic lupus erythematosus, Sjögren syndrome, and systemic sclerosis. An algorithm including genetic markers, serological markers, and clinical data will provide information to the patient and the care provider. This will help in early diagnosis of an existing disease as well as in estimating the risk for an emerging rheumatic disease in individuals with symptoms such as joint pain but without current signs of inflammation. As proposed by Weyrich et al [37] this triage system is user-friendly, dynamic, and incorporates the potential of eHealth [38]. A recent review [38] identified a lack of mHealth and eHealth tools in the field of rheumatology, underlining the innovative and stand-alone character of the JPAST project. In the first step, the diagnostic accuracy of JPAST and its usability will be analyzed in patients newly presenting to secondary care–based rheumatology clinics to test its usefulness. JPAST performance can then be compared to currently used local screening methods [17], which will allow further algorithm improvement. Once an acceptable level of performance is achieved, a prospective primary care–based long-term study should compare the clinical and economic impact of using this system versus current local care [39]. We believe JPAST will provide an accelerated pathway and improved personalized diagnosis of autoimmune inflammatory RMDs by combining innovative products and services and including all main stakeholders.

## Acknowledgments

The authors thank all partners and collaborators for their support of this project. JPAST is a public–private partnership program for the digitalization of health care, supported by EIT Health. This project has received funding from EIT Health. EIT Health is supported by the European Institute of Innovation and Technology (EIT), a body of the European Union that receives support from the European Union’s Horizon 2020 Research and Innovation program. Thermo Fisher Scientific is the participating

diagnostic company developing the serology and genetics tools. Scientists/clinicians from the four contributing and coauthoring centers have developed the eHealth tool “Rheumatic?” with help from the strategic design agency Ocean.

## Authors' Contributions

All authors wrote, reviewed, and approved the final manuscript. Further members of the JPAST Group are as follows: Monika Hansson (Division of Rheumatology, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden), Sascha Swiniarski (Thermo Fisher Scientific, Freiburg, Germany), Leonid Padyukov (Division of Rheumatology, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden), Caroline Grönwall (Division of Rheumatology, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden), Tom Huizinga (Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands), Andrew Filer (Institute of Inflammation and Ageing, University of Birmingham, Birmingham, UK), Sofia Svanteson (Ocean Observations, Stockholm, Sweden), Alexandra Lindfors (Ocean Observations, Stockholm, Sweden), and Aase Hensvold (Division of Rheumatology, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden).

## Conflicts of Interest

LK is a board member of the Riskminder Foundation, which has ownership in the eHealth company Elsa, which aims to continue the development of certain aspects of the JPAST tools. EE, IG, LMA, and MP work for Thermo Fisher Scientific. TB works for Ocean Observations. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Screencast of "Rheumatic?" the JPAST (Joint Pain Assessment Scoring Tool) symptom checker.
[MOV File , 22792 KB - mhealth_v8i5e17507_app1.mov ]

## References


1. Quinn MA, Emery P. Window of opportunity in early rheumatoid arthritis: possibility of altering the disease process with early intervention. Clin Exp Rheumatol 2003;21(5 Suppl 31):S154-S157. [Medline: 14969068]
2. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. Lancet 2016 Oct 22;388(10055):2023-2038. [doi: 10.1016/S0140-6736(16)30173-8] [Medline: 27156434]
3. Combe B, Landewe R, Daien CI, Hua C, Aletaha D, Álvaro-Gracia JM, et al. 2016 update of the EULAR recommendations for the management of early arthritis. Ann Rheum Dis 2016 Dec 15;76(6):948-959. [doi: 10.1136/annrheumdis-2016-210602]
4. Hider SL, Muller S, Helliwell T, Prior JA, Scott I, Lawton SA, et al. Symptoms associated with inflammatory arthritis are common in the primary care population: results from the joint symptoms survey. Rheumatology (Oxford) 2019 Nov 01;58(11):2009-2014. [doi: 10.1093/rheumatology/kez184] [Medline: 31079150]
5. Raza K, Stack R, Kumar K, Filer A, Detert J, Bastian H, et al. Delays in assessment of patients with rheumatoid arthritis: variations across Europe. Ann Rheum Dis 2011 Aug 07;70(10):1822-1825. [doi: 10.1136/ard.2011.151902]
6. Simons G, Belcher J, Morton C, Kumar K, Falahee M, Mallen CD, et al. Symptom Recognition and Perceived Urgency of Help-Seeking for Rheumatoid Arthritis and Other Diseases in the General Public: A Mixed Method Approach. Arthritis Care & Research 2017 Apr 10;69(5):633-641. [doi: 10.1002/acr.22979]
7. Simons G, Lumley S, Falahee M, Kumar K, Mallen CD, Stack RJ, et al. The pathway to consultation for rheumatoid arthritis: exploring anticipated actions between the onset of symptoms and face-to-face encounter with a healthcare professional. BMC Musculoskelet Disord 2017 Jun 14;18(1). [doi: 10.1186/s12891-017-1619-9]
8. Kumar K, Daley E, Carruthers DM, Situnayake D, Gordon C, Grindulis K, et al. Delay in presentation to primary care physicians is the main reason why patients with rheumatoid arthritis are seen late by rheumatologists. Rheumatology 2007 Jun 27;46(9):1438-1440. [doi: 10.1093/rheumatology/kem130]
9. Raciborski F, Kłak A, Kwiatkowska B, Batko B, Sochocka-Bykowska M, Zoń-Giebel A, working team. Diagnostic delays in rheumatic diseases with associated arthritis. Reumatologia 2017;55(4):169-176 [FREE Full text] [doi: 10.5114/reum.2017.69777] [Medline: 29056771]
10. Stack RJ, Nightingale P, Jinks C, Shaw K, Herron-Marx S, Horne R, et al. Delays between the onset of symptoms and first rheumatology consultation in patients with rheumatoid arthritis in the UK: an observational study. BMJ Open 2019 Mar 04;9(3):e024361. [doi: 10.1136/bmjopen-2018-024361]
11. van der Linden MPM, le Cessie S, Raza K, van der Woude D, Knevel R, Huizinga TWJ, et al. Long-term impact of delay in assessment of patients with early arthritis. Arthritis & Rheumatism 2010 Nov 30;62(12):3537-3546. [doi: 10.1002/art.27692]
12. Zink A, Schneider M. Versorgungsplanung in der Rheumatologie: Eckpunkte aus dem 2016 aktualisierten Memorandum der Deutschen Gesellschaft für Rheumatologie zur Versorgungsqualität. Akt Rheumatol 2018 Mar 21;43(05):390-394. [doi: 10.1055/a-0573-8722]

13. Ehrenstein B, Pongratz G, Fleck M, Hartung W. The ability of rheumatologists blinded to prior workup to diagnose rheumatoid arthritis only by clinical assessment: a cross-sectional study. Rheumatology (Oxford) 2018 Sep 01;57(9):1592-1601. [doi: 10.1093/rheumatology/key127] [Medline: 29850885]
14. March L, Smith EU, Hoy DG, Cross MJ, Sanchez-Riera L, Blyth F, et al. Burden of disability due to musculoskeletal (MSK) disorders. Best Practice & Research Clinical Rheumatology 2014 Jun;28(3):353-366. [doi: 10.1016/j.berh.2014.08.002]
15. Feuchtenberger M, Nigg AP, Kraus MR, Schäfer A. Rate of Proven Rheumatic Diseases in a Large Collective of Referrals to an Outpatient Rheumatology Clinic under Routine Conditions. Clin Med Insights Arthritis Musculoskelet Disord 2016 Oct 02;9:181-187. [doi: 10.4137/cmamd.s40361]
16. Villeneuve E, Nam JL, Bell MJ, Deighton CM, Felson DT, Hazes JM, et al. A systematic literature review of strategies promoting early referral and reducing delays in the diagnosis and management of inflammatory arthritis. Ann Rheum Dis 2012 Apr 24;72(1):13-22. [doi: 10.1136/annrheumdis-2011-201063]
17. Benesova K, Lorenz H, Lion V, Voigt A, Krause A, Sander O, et al. Früh- und Screeningsprechstunden: Ein notwendiger Weg zur besseren Frühversorgung in der internistischen Rheumatologie? Z Rheumatol 2019 Aug 29;78(8):722-742. [doi: 10.1007/s00393-019-0683-y]
18. Krey J. "To triage or not to triage" … (frei nach Shakespeare). Med Klin Intensivmed Notfmed 2016 Jul 18;111(6):565-566. [doi: 10.1007/s00063-016-0200-x]
19. Thompson AE, Haig SL, LeRiche NG, Rohekar G, Rohekar S, Pope JE. Comprehensive Arthritis Referral Study — Phase 2: Analysis of the Comprehensive Arthritis Referral Tool. J Rheumatol 2014 Sep 01;41(10):1980-1989. [doi: 10.3899/jrheum.140167]
20. Gerdtz MF, Chu M, Collins M, Considine J, Crellin D, Sands N, et al. Factors influencing consistency of triage using the Australasian Triage Scale: implications for guideline development. Emerg Med Australas 2009 Aug;21(4):277-285. [doi: 10.1111/j.1742-6723.2009.01197.x] [Medline: 19682012]
21. Powley L, McIlroy G, Simons G, Raza K. Are online symptoms checkers useful for patients with inflammatory arthritis? BMC Musculoskelet Disord 2016 Dec 24;17(1):362 [FREE Full text] [doi: 10.1186/s12891-016-1189-2] [Medline: 27553253]
22. El-Kareh R, Hasan O, Schiff GD. Use of health information technology to reduce diagnostic errors. BMJ Qual Saf 2013 Oct;22 Suppl 2:ii40-ii51 [FREE Full text] [doi: 10.1136/bmjqs-2013-001884] [Medline: 23852973]
23. Garg AX, Adhikari NKJ, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. JAMA 2005 Mar 9;293(10):1223-1238. [doi: 10.1001/jama.293.10.1223] [Medline: 15755945]
24. Riches N, Panagioti M, Alam R, Cheraghi-Sohi S, Campbell S, Esmail A, et al. The Effectiveness of Electronic Differential Diagnoses (DDX) Generators: A Systematic Review and Meta-Analysis. PLoS One 2016;11(3):e0148991 [FREE Full text] [doi: 10.1371/journal.pone.0148991] [Medline: 26954234]
25. Ronicke S, Hirsch MC, Türk E, Larionov K, Tientcheu D, Wagner AD. Can a decision support system accelerate rare disease diagnosis? Evaluating the potential impact of Ada DX in a retrospective study. Orphanet J Rare Dis 2019 Dec 21;14(1):69 [FREE Full text] [doi: 10.1186/s13023-019-1040-6] [Medline: 30898118]
26. Semigran HL, Linder JA, Gidengil C, Mehrotra A. Evaluation of symptom checkers for self diagnosis and triage: audit study. BMJ 2015;351:h3480 [FREE Full text] [Medline: 26157077]
27. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO, et al. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis Rheum 2010 Sep;62(9):2569-2581 [FREE Full text] [doi: 10.1002/art.27584] [Medline: 20872595]
28. Aringer M, Costenbader KH, Dörner T, Johnson SR. Limited role of specificity of ANA for SLE classification. Response to: 'European League against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus: the laboratory immunologist's point of view' by Infantino. Ann Rheum Dis 2019 Dec 12;78(9):1151. [doi: 10.1136/annrheumdis-2019-216700]
29. van den Hoogen F, Khanna D, Fransen J, Johnson S, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis 2013 Nov;72(11):1747-1755. [doi: 10.1136/annrheumdis-2013-204424] [Medline: 24092682]
30. Lundberg IE, Tjärnlund A, Bottai M, Werth VP, Pilkington C, Visser M, International Myositis Classification Criteria Project consortium, The Euromyositis register, The Juvenile Dermatomyositis Cohort Biomarker Study and Repository (JDRG) (UKIreland). 2017 European League Against Rheumatism/American College of Rheumatology classification criteria for adult and juvenile idiopathic inflammatory myopathies and their major subgroups. Ann Rheum Dis 2017 Dec;76(12):1955-1964 [FREE Full text] [doi: 10.1136/annrheumdis-2017-211468] [Medline: 29079590]
31. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjögren's syndrome. Ann Rheum Dis 2016 Oct 26;76(1):9-16. [doi: 10.1136/annrheumdis-2016-210571]
32. Rantapää-Dahlqvist S, de Jong BAW, Berglin E, Hallmans G, Wadell G, Stenlund H, et al. Antibodies against cyclic citrullinated peptide and IgA rheumatoid factor predict the development of rheumatoid arthritis. Arthritis & Rheumatism 2003 Oct 07;48(10):2741-2749. [doi: 10.1002/art.11223]

33. Svendsen AJ, Kyvik KO, Houen G, Junker P, Christensen K, Christiansen L, et al. On the Origin of Rheumatoid Arthritis: The Impact of Environment and Genes—A Population Based Twin Study. PLoS ONE 2013 Feb 28;8(2):e57304. [doi: 10.1371/journal.pone.0057304]
34. Eroglu GE, Kohler PF. Familial systemic lupus erythematosus: the role of genetic and environmental factors. Ann Rheum Dis 2002 Jan;61(1):29-31 [FREE Full text] [doi: 10.1136/ard.61.1.29] [Medline: 11779754]
35. Drouin A, Letarte G, Raymond F, Marchand M, Corbeil J, Laviolette F. Interpretable genotype-to-phenotype classifiers with performance guarantees. Sci Rep 2019 Mar 11;9(1). [doi: 10.1038/s41598-019-40561-2]
36. Thrall JH. Data Mining, Knowledge Creation, and Work Process Enhancement in the Second Generation of Radiology's Digital Age. Journal of the American College of Radiology 2013 Mar;10(3):161-162. [doi: 10.1016/j.jacr.2012.08.007]
37. Weyrich P, Christ M, Celebi N, Riessen R. Triagesysteme in der Notaufnahme. Med Klin Intensivmed Notfmed 2012 Feb 1;107(1):67-79. [doi: 10.1007/s00063-011-0075-9]
38. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR mHealth uHealth 2019 Aug 05;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
39. Schwarting A, Dreher M, Assmann G, Witte T, Hoeper K, Schmidt RE. Erfahrungen und Ergebnisse aus Rheuma-VOR. Z Rheumatol 2019 Aug 29;78(8):743-752. [doi: 10.1007/s00393-019-00694-1]


## Abbreviations

**ACPA:** anticitrullinated protein antibody
**ACR:** American College of Rheumatology
**DDSS:** diagnostic decision support system
**eHealth:** electronic health
**EIT:** European Institute of Innovation and Technology
**EU:** European Union
**EULAR:** European League Against Rheumatism
**GP:** general practitioner
**GWAS:** genome-wide association studies
**JPAST:** Joint Pain Assessment Scoring Tool
**NHS:** National Health Service
**RA:** rheumatoid arthritis
**RF:** rheumatoid factor
**RMD:** rheumatic and musculoskeletal diseases
**SLE:** systemic lupus erythematosus
**SS:** Sjögren syndrome
**SSc:** systemic sclerosis


*Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by M Wong, B Eapen, F Dittrich; comments to author 31.01.20; revised version received 10.02.20; accepted 22.03.20; published 15.05.20.*

*Please cite as:*
*Knitza J, Knevel R, Raza K, Bruce T, Eimer E, Gehring I, Mathsson-Alm L, Poorafshar M, Hueber AJ, Schett G, Johannesson M, Catrina A, Klareskog L, JPAST Group*
*Toward Earlier Diagnosis Using Combined eHealth Tools in Rheumatology: The Joint Pain Assessment Scoring Tool (JPAST) Project*
*JMIR Mhealth Uhealth 2020;8(5):e17507*
*URL: https://mhealth.jmir.org/2020/5/e17507*
*doi:10.2196/17507*
*PMID:32348258*

Viewpoint

# Understanding and Preventing Health Concerns About Emerging Mobile Health Technologies

Frank T Materia[1], MS, MHS; Kate Faasse[2], DPhil; Joshua M Smyth[1], DPhil

[1]The Pennsylvania State University, University Park, PA, United States

[2]University of New South Wales, Sydney, Australia

**Corresponding Author:**
Joshua M Smyth, DPhil
The Pennsylvania State University
219 Biobehavioral Health Building
University Park, PA, 16802
United States
Phone: 1 814 863 8402
Email: Smyth@psu.edu

## Abstract

New technologies and innovations have often improved population well-being and societal function; however, these are also often initially accompanied by worry and fear. In some cases, such worries can impede, or even prevent entirely, the adoption of the technology. Mobile health (mHealth), a discipline broadly focused on employing ambulatory technologies to improve the affordability, reach, and effectiveness of health promotion and clinical intervention approaches, offers new innovations and opportunities. Despite emerging evidence supporting mHealth efficacy (eg, for improving health outcomes), some individuals have concerns about mHealth technology that may impede scalability, efficacy, and, ultimately, the public health benefits of mHealth. We present a review and conceptual framework to examine these issues, focusing on three overarching themes: biophysiological, psychological, and societal concerns. There are features of mHealth that lead to worries about the potential negative effects on an individual's health (eg, due to exposure to electromagnetic or radio waves), despite evidence supporting the safety of these technologies. When present, such beliefs can lead to worry that gives rise to the experience of unpleasant and concerning physical symptoms—the nocebo effect. This may represent an important implementational barrier because of apprehension toward beneficial mHealth products (or features thereof, such as wireless charging, wearable or implantable sensors, etc) and may also have broader ramifications (eg, leading to economic, governmental, and legislative actions). In addition to reviewing evidence on these points, we provide a broad three-step model of implementation research in mHealth that focuses on understanding and preventing health concerns to facilitate the safe and effective scalability of mHealth (and that may be generalizable and applied to similar technologies): (1) evaluating and better discerning public perceptions and misperceptions (and how these may differ between populations), (2) developing theory-based public health communication strategies regarding the safety of mHealth, and (3) disseminating this messaging using evidence-based methods. Collectively, these steps converge on reviewing evidence regarding the potential role of worry and nocebo in mHealth and providing a model for understanding and changing attitudes and preventing unfounded negative perceptions related to mHealth technology.

***(JMIR Mhealth Uhealth 2020;8(5):e14375)*** doi:10.2196/14375

**KEYWORDS**

mHealth; technology; nocebo effect; implementation science; medically unexplained symptoms

## Introduction

Human technological innovation has progressed tremendously over the past century. From the flagship electronic devices of the early 1900s to the advanced circuits of the midcentury, the proliferation of the personal computer and the internet in the 1980s and 1990s, and the current expansion of nanotechnology and smart devices, the rapid evolution of technology presents opportunities for improving societal function and global well-being. One particular scientific field, *mobile health* (mHealth), is an example of technological expansion aimed at improving human outcomes.

Over the past 20 years, mHealth has emerged as an integrative discipline, focusing on developing and implementing wireless, portable, or implantable technology for improving human health [1-3]. Some mHealth approaches such as wearable fitness

trackers for encouraging physical activity and smartphone apps supporting medication adherence have become ubiquitous in modern society as tools for promoting health [4]. Infrastructures across the world, at multiple levels, continue to adapt to support the introduction of mHealth devices to the masses. Some examples include investments in expanding cellular infrastructure and internet access [5], advertising campaigns disseminating new mHealth products [6], collaborations between engineers and physicians for developing innovative mobile intervention approaches [7], and electronic systems integrating mobile technology into clinical treatment approaches within health care organizations [8,9]

## Mobile Health Technology's Potential for Improving Health

The rapid proliferation of new technologies has often outpaced the slower process of collecting evidence on mHealth, such as by conducting rigorous clinical trials [1]. However, there is a growing scientific evidence base supporting mHealth's effectiveness in improving health outcomes. Recently published clinical trial results, accompanied by evidence derived from alternative research methodologies (eg, microrandomized trials and *big-data* analytic approaches), increasingly support the efficacy [2] and safety [10] of mHealth approaches for improving various health behaviors and outcomes (eg, smoking cessation [11], HIV care [12], medication adherence [13], chronic disease management and care [2], and health-related quality of life [14]) and reducing traditional care costs [15].

## Need for Human-Centered Implementation Research

Even with emerging evidence suggesting the benefits of mHealth, the expansion of such technologies across multiple levels of society worldwide calls for *implementation* research—investigations aimed at better understanding the factors related to the successful introduction and utilization of mHealth interventions in the real-world community and clinical settings [16]. Although limited, some mHealth implementation research does exist, and there are frameworks developed (eg, the Integrate, Design, Assess, and Share – IDEAS Framework [17]) for digital interventions that broadly suggest methods for designing, implementing, and disseminating such products. Overall, however, the literature regarding best practices for implementing mHealth is unclear. There is some consensus that one of the most crucial prescriptions for advancing mHealth implementation research involves better understanding the human-centered factors associated with the adoption, uptake, and sustained use of new technologies [18]. Even if mHealth systems are shown to be effective, if individuals are unwilling to accept, trust, and engage with such technologies, mHealth cannot reach its potential for improving public health.

## Issues in Widespread Technology Adoption

Given there is no *gold standard* for conducting human-centered mHealth implementation science [19], historical perspectives regarding mass technology introductions may provide an evidence-informed precedent for approaching such research questions. Large-scale technological implementations of the past century provide exemplar accounts of human-centered factors that may hinder the widespread implementation and utilization of mHealth. For example, unsupported health worries surrounding the safety of microwave ovens (ie, concerns about microwave radiation being emitted into the nearby environment, increasing a perceived risk of cancer) date back to the product's introduction in the 1940s [20], and these worries were further exacerbated in the 1960s when a US government survey reinforced this (mis)perception [21]. The expansion of mobile phones and cellular networks in the 1980s brought about public trust issues with the increased amount of unseen radio waves traveling through the air [22]. Recently, public concerns have been raised about the use of wind turbines for sustainable energy and almost unavoidable exposure to Wi-Fi internet signals and related to the upcoming global transition to 5G cellular data connectivity [23-27]. These factors tend to revolve around worries and distrust about the effects that unfamiliar, innovative devices may have on physical well-being [28]. Although many people support such advancements, some are hesitant and express concerns about potential risks—particularly about negative health effects—of new technologies [29].

In addition, the recognition that large corporate entities (eg, tech companies, banks, and communication providers) increasingly capture large amounts of data for their own use—often without clear disclosure and, in some cases, without consent—has stoked concerns about data privacy and the usage of such data [30,31]. These data concerns, coupled with existent worries about the risks of new technologies on physical health, have led to distrust and may ultimately culminate in technological avoidance behaviors [32].

## Biopsychosocial Approach to Understanding Mobile Health Concerns

We propose that worries about new technologies are multifaceted and that implementational concerns can be characterized as stemming from the dynamic interplay of biophysiological (eg, concerns about mHealth affecting physical health), psychobehavioral (eg, affecting well-being, perceptions, and decision making), and social (eg, affecting larger global policy and regulatory practices) constructs. Given mHealth's integration and reliance upon varied technological platforms (eg, passive sensing, wearable components, near-field communication [NFC] transmitters), we view mHealth technology as a case study of how current widespread implementation of innovative devices (with great potential for improving human health) may inadvertently instigate public distrust and concern. These issues likely are not confined to mHealth; many related technological services and products (eg, wireless power transfer, smart homes, and Wi-Fi connectivity)

serve other market interests beyond health promotion and intervention yet may share features that similarly elicit concerns or distrust. We thus hope our attempt to better understand and prevent biopsychosocial public concerns with mHealth may help inform related technologies and issues.

## Objectives

The overarching objective of this paper is to bring attention to the potential biophysiological, psychological/behavioral, and societal ramifications related to (often unfounded) health concerns with newly emergent mHealth technologies. Given that there is very limited existing literature about this phenomenon, it is not yet possible to conduct a systematic review on this topic. Rather, we aimed to narratively summarize the key issues, present selected evidence, and, more generally, provide a conceptual framework for these issues—particularly for readers who develop novel mHealth technology (including those outside of social and behavioral science, such as engineers). Notably, we posit that there already likely are—and will continue to be—potential implementation barriers in the rapidly emergent mHealth product market. Such barriers, if not evaluated and addressed, may lead to lower uptake and nonadherence with mHealth technology. Specifically, we highlight the potential role of concerns regarding *power* (eg, electromagnetic hypersensitivity syndrome, EHS) and the nocebo effect, consumer distrust and technology avoidance behaviors, and larger social implications for user acceptability, regulation, and market availability. We also attempted to indicate how such considerations may extend to other (similar) technologies besides mHealth.

The absence of empirical evidence in this context also means the prevalence and magnitude of mHealth worries and concern are, at this point, somewhat left to speculation. To address this dearth in the literature, we believe it is essential to outline a conceptual framework to stimulate and inform needed research. Therefore, a secondary goal is to propose an approach to studying the details of such mHealth-related worries and concerns and suggest public health prevention efforts for curbing unfounded distrust in mHealth technologies. The proposed framework also integrates elements from widely accepted models for implementation research (eg, the Reach, Effectiveness, Adoption, Implementation and Maintenance-RE-AIM Framework [33] and the Consolidated Framework for Implementation Research – CFIR [34])—broadly speaking, evaluating the target population's perceptions, designing interventions to match needs, and disseminating materials with evidence-informed methodologies. Future directions for evaluating and enhancing the uptake of safe and effective mHealth technologies are presented.

## Biophysiological Concerns With Mobile Health Implementation

Worries about how novel technologies influence health outcomes are prevalent and appear to be increasing with the proliferation of advanced modern technology [29]. Many new technologies, just as have many *old* technologies over the past century, have been linked to public concerns and reports of unpleasant physical symptoms and adverse health outcomes that sufferers attribute to exposure to new technologies [35]. Symptoms reported are typically not associated with a specific bodily system or an underlying disease process but, nonetheless, are reported as severe and disabling; these include, for example, sleep disturbances, fatigue, headaches, and cognitive problems [35]. Such concerns about the potential harms of new technologies and other aspects of modern life are common among otherwise healthy individuals [36]. Public concerns appear to be heightened by technological exposures that have certain characteristics, for example, exposures that are involuntary, inescapable, and from a novel or unfamiliar source [37].

### Electromagnetic Hypersensitivity Syndrome and Mobile Health

One feature common to many technologies, including many mHealth devices, that prompts public concerns about health impacts is the emission of electromagnetic radiation. A number of products and components necessary for supporting mHealth systems (eg, Wi-Fi routers, cellular towers, rechargeable batteries, and wireless data) rely on power to operate harness magnets, capture or emit radio waves, or send or receive electricity. Some individuals have concerns about the release of unsafe levels of electromagnetic radiation into the environment. Furthermore, it is purported that there are specific types of predisposed or particularly sensitive individuals who experience unpleasant symptoms following such perceived exposures. Such people are often described as having EHS [35]. Some studies have estimated that around 1 in 20 people in the general population believe they are affected by EHS, and it appears that many more are concerned about the potential negative health effects of exposure to electromagnetic radiation [38,39].

It is likely that media coverage can strongly shape individuals' perceptions toward potentially inaccurate (and sometimes risky) misinformation (eg, the antivaccination movement [40] and dangerous diet fads [41]). In recent years, the reporting of EHS in the media has increased public concerns and subsequently reported instances and symptoms, as attributed to electromagnetic frequencies in the environment [42]. Such circumstances have led to general public distress and distrust, increased seeking of medical care for EHS symptoms, and a growing number of activists and community groups worldwide [24]. Some EHS sufferers and their families have filed lawsuits to get rid of technology that they view as posing risks to environment (including those vital to mHealth, such as Wi-Fi), asserting that such exposures are the cause of negative physical and psychological health outcomes, including suicidal ideation [43]. Other EHS sufferers have taken more extreme intervention measures, such as choosing to leave their families and re-establish their lives in desolate rural areas where cellular networks are nonexistent [24]. We recognize that the veracity of claims from media outlets can be questionable; what these reports do suggest, however, is that concerns and worries related to technology appear prominent, broadly distributed (eg, geographically and demographically), and accompanied by

behavioral responses intended to manage the concerns and worries.

### Science Behind Electromagnetic Hypersensitivity Syndrome

In response to rapidly growing public concerns, observational and experimental studies have been conducted to understand the prevalence and possible physiological mechanisms of EHS. Cross-sectional research has found that the majority of EHS sufferers are self-diagnosed and have not sought medical care for the condition, that the mean duration of EHS symptoms is 10.5 years, and that more than half of those with EHS actively try to avoid electromagnetic field (EMF) sources [44]. Rigorous trials (both in the lab and out in the field) over the past 20 years have also been conducted to better understand EHS and the effects of electromagnetic radiation on human health. The general consensus of the scientific and medical communities, based on extensive evidence, including that from double-blind sham-controlled experimental studies, is that the symptoms reported by individuals with EHS are *not* caused by exposure to electromagnetic radiation. When people believe that they are (or might be) exposed to EMF, they report experiencing unpleasant EHS symptoms, regardless of whether the EMF-emitting devices are switched on or off [35,45-48]. Multiple systematic reviews have found that there is no evidence to support EHS as a disease with a biological basis [49,50] and that there has never been a causal relationship established between exposures to electromagnetic radiation and well-being [51]. An alternative viewpoint that EHS symptoms are likely to be caused by the expectations and beliefs about the risk of potential harm and the accompanying worry is emerging; this process has more broadly been described as the nocebo effect.

### Nocebo Effect

This is a phenomenon whereby nonspecific, unpleasant physical symptoms and other adverse health outcomes occur in response to negative expectations and beliefs (in this case, related to exposure to electromagnetic radiation)—rather than any specific physiological or environmental cause [28,52-54]. Nocebo effects are prevalent in medicine and daily life [55], including in the development and maintenance of EHS symptoms, and these effects can be powerful enough to cause symptoms that lead individuals to hospitalization and medical intervention [56]. Although the empirical evidence shows that exposure to multiple types of invisible wave frequencies does not cause EHS, individuals are still reporting unpleasant symptoms and negative outcomes as a result of (real or perceived) exposure. As noted, the (likely mistaken) perception that EMF exposure is unsafe leads to experiencing associated concerns, negative expectations, and subsequent (nocebo) symptoms.

Negative expectations can act via two different pathways to cause nocebo effects. First, by directly influencing the experience of physical symptoms through neurobiological changes as well as increasing anxiety and shifting attention toward the expected experience [57]. Second, via a misattribution process whereby normal physical symptoms are mistakenly attributed to the perceived toxic or harmful exposure (ie, in this case, electromagnetic radiation). The experience of physical symptoms is common, even in healthy individuals, thereby providing a range of potential symptoms for misattribution at any given time [58].

### Mobile Health Innovations May Fuel Nocebo Effects

Given the possibility that negative expectations regarding the perceived risk resulting from exposure to electromagnetic radiation can fuel nocebo symptoms and the common use of EHS-relevant technologies in the mHealth field (eg, Wi-Fi, wireless charging, Bluetooth connectivity, cellular signals, etc), evaluating and addressing worries about mHealth technologies is timely. Technologies that are poorly understood by the population at large are increasingly being employed to support mHealth prevention and treatment approaches. For example, wireless power transfer is now necessary for charging current health-monitoring smartwatches and wearables [59]; in addition, implantable pacemakers rely on this technology for seamless charging without surgery [3]. NFC devices, which can be used to trigger intervention messaging in particular spaces and environments on mobile devices, rely upon the sending and receipt of radio waves [60]. As these technologies continue to proliferate to support mHealth approaches, the likelihood of such devices fueling health concerns, negative expectations, and EHS symptoms via the nocebo effect is heightened. More generally, assuming that mHealth technologies and therapies are developed that are, in fact, safe and effective, these worries represent a critical implementational barrier to the dissemination and impact of mHealth. Thus, research on these topics may also be helpful to understand and address public distrust, thereby facilitating the expansion of safe and effective mHealth technologies (and providing the opportunity for such technologies to reach their potential for improving health).

## Psychological/Behavioral Concerns With Mobile Health Implementation

Roger's [61] *Diffusion of Innovations Model* of the 1960s (among others) aimed to characterize individuals across a spectrum of psychological and behavioral dispositions toward adopting new technologies, from *early adopter* to *laggard*. However, given recent surges in the accessibility of mobile technologies, and that over 90% of global adults now own a mobile device [62], the many *mobile laggards* of recent decades have become potential mHealth consumers/users. The majority of people worldwide are surrounded by new technological systems and infrastructure (eg, video surveillance, open Wi-Fi, and long-term evolution networking) [5]. Separate from contributing to EHS and nocebo effects, the global push to *go mobile* has raised significant public concerns about data privacy and trust issues.

One-fourth of the world population logs on to Facebook at least once per month [63], and almost 88% of the global smartphone marketplace belongs to Google [64]. In recent years, such companies (among many others, such as Equifax [65] and Amazon [66]) have faced public scrutiny over their mass procurement, guarding, and utilization of personal and mobile usage data [67]. Much of these data are collected *passively*, as in, users are unaware of when their personal mobile device, browsing, or provided data are being collected, or how they are

being harnessed (or used) by the associated company [68]. Many consumers, especially in high-income nations, view the procurement of ongoing, passively obtained personal, financial, and health information to be intrusive and an invasion of privacy [68,69].

Many tech giants (eg, Facebook and Google) assert that users fully accept their companies' data privacy standards when they agree to the *terms and conditions* of signing up for a service [70]. Although such affirmation has in some cases been seen as legally binding [71], such data privacy considerations have also been widely criticized. mHealth technologies, particularly those that involve personal health information (PHI), may require greater attention to these data privacy concerns; this may result in better data management practices, but this also represents a potential barrier to implementation and adoption of mHealth technologies [72]. For example, clinically verified monitors for tracking ambulatory physiology (eg, blood pressure) face stringent US Food and Drug Administration standards for the transfer and storage of data to protect patient anonymity [73], and similar standards exist in other high-income nations [74]. Many private companies that manufacture medical-grade mHealth devices have also made their data privacy actions transparent on the Web, many aspiring to use layperson language [75-77]. Physicians have also requested clinician-focused educational materials about the security of mHealth technology's data transfer and privacy, and easily delivered resources for ensuring enhanced transparency (to patients) about the safety of implementing new technologies into care [6]. Yet given the public distrust fueled by concerns with the tech giants' usage of individuals' mobile data, trust in upcoming mHealth technologies may dwindle—if so, such worries and concerns may slow the adoption and uptake of such devices. Indeed, researchers have noted considerable variability in perceived safety and security of mHealth apps and similar products [31].

## Mobile Health Innovations May Fuel Avoidance Behaviors

Avoidance behaviors are characterized by protective, sometimes proactive, actions toward avoiding situations that may be perceived as harmful to well-being [78]. In terms of innovative technologies, Technology Threat Avoidance Theory describes consumers' actions for avoiding technological entities that are perceived to cause harmful outcomes, especially individual harms related to information and data security [32,79]. Although many mHealth technologies have secure collection and responsible usage of patient data for clinical purposes [7], broader concerns regarding general consumer security (eg, of big data procurement) are likely to persist even as mHealth continues to be more integrated into clinical prevention and intervention approaches—perhaps resulting in increased avoidance behaviors.

Worries regarding *data insecurity* [67] are certainly not constrained to mHealth. For example, SmartHome technologies (ie, wireless monitoring of heating and air conditioning, locks, and surveillance systems in homes, among other capabilities) are increasingly prevalent in high-income nations and require the transfer of intimate home-based environmental data and video streams between mobile devices [80]. In addition, social media utilization increased by 13% globally from 2017 to 2018, with most platforms being accessed via mobile devices [81], and the background monitoring of smartphone usage statistics and *digital footprints* is at an all-time high [82]. Multiple private and public technology companies, including those invested in mHealth and other consumer innovations, will benefit from a more nuanced and evidence-based understanding and accurate evaluation of perceived data security (or lack thereof) and its impact on avoidance behaviors.

## Societal Concerns With Mobile Health Implementation

For mHealth technologies to reach their full potential on a global level, they must be trusted, accepted, and adopted by society at large. One implementation factor positively associated with the adoption of newly innovative technologies to society is overall consumer *acceptability*. Multiple well-regarded models have been established to explicate the relationship between consumer acceptability and adoption of new technologies, including the Technology Acceptance Model [83] and the Consumer Acceptance of Technology model [84]. Acceptability is purported to be a multifaceted concept, with agreeability, satisfaction, and willingness to engage with a device being common constructs associated with measuring how *acceptable* a product is to an end user [19]. It can also be measured at different timepoints, including prior to use, to predict adherence with a new technology (eg, *hypothetical*, *a priori*, or *prospective* acceptability) and after (eg, *actual*, *experienced*, or *retrospective* acceptability) the technology has been introduced to the user [85,86]. Higher levels of consumer acceptability (both before and after engaging with a device) are associated with higher sustained levels of uptake to using technological systems [83,84].

The concern here is that for mHealth, we have strong reasons to believe that some significant fraction of potential consumers may have worries regarding potential negative health effects and data security and privacy, particularly as they see (likely nonrepresentative, but very powerful) media stories consistent with their concerns. In such cases, it appears likely that this will limit consumers' acceptance of innovative mHealth approaches, their willingness to engage with mHealth, and their satisfaction with the technology (ie, devices thought to elicit negative health effects or foster data privacy concerns will not likely be satisfactory to the average consumer). Taken together, these effects will limit consumer mHealth acceptability, which, in turn, limits social uptake of the technology overall. If the individuals who can benefit from mHealth do not accept the technology, particularly when based on unfounded concerns, this may impede the larger societal acceptance and capacity of mHealth to reach its potential for public good as a tool for enhancing human health.

## Issues for Taking Mobile Health to Market

Government agencies tasked with ensuring public safety and health are likely to pay attention to any data privacy concerns or perceived negative health effects of mHealth products, even in the absence of a strong scientific base to support such worries. Addressing these concerns can be costly. For example, investigating complaints about wind farm–related EHS is anticipated to cost the Australian government more than Aus $2 million (US $1.2 million dollars) [23]. Chapman and Crichton [23] note that unfounded public concerns not only cause major delays in the rollout of beneficial technology but also impede the pace of global technological advancements. A growing rollout of mHealth to market in high-income nations in the near future may face similar scrutiny, potentially requiring taxpayer dollars, reducing government interest in supporting such technologies, and ultimately slowing the development and implementation of mHealth.

In addition, reported concerns with mHealth technologies can result in legal ramifications for product developers. For example, in the case of EHS, some sufferers have filed lawsuits against producers and local governing bodies in an attempt to rid their environments of various types of devices that emit electromagnetic radiation [43]. As similar public perceptions could evolve around mHealth, manufacturers who would otherwise be well fit to mass produce such technologies may be hesitant to embark on such an endeavor for fear of litigation, negative publicity, and lawsuits. Alternatively, if such concerns do emerge about a recently marketed product, the resultant legal and financial challenges may be beyond the capacity of the business to manage. In addition, extending prior discussion on this point, in recent years, tech giants such as Facebook and Apple have faced class-action legal suits regarding data privacy issues, in turn, contributing to publicized distrust and worry with associated apps and products [68,87].

As a function of reduced acceptability and uptake, the mHealth product market may be limited or may fail to show the expected growth. As has happened in the past when unfounded privacy and health concerns have been disseminated to the public, otherwise safe products are seen as unacceptable and these new technologies are avoided by consumers [23,83,84]. This, in turn, limits sales, meaning producers suffer low return on research and development expenditures. Subsequently, product developers halt manufacturing, shifting their resources and interests toward more acceptable products. The compounded scenario of distrust in mobile technologies, low acceptability limiting social capital and market reach, and hence, low producer investment in developing such devices, all reduce the likelihood of mHealth reaching its full potential for enhancing public health.

## Social and Behavioral Science Solutions to Address Mobile Health Implementation

To better understand and recalibrate public concerns surrounding emerging mHealth technologies, we propose a biopsychosocial framework derived from previous work to investigate these human-centered implementation factors. Such research can inform proactive prevention efforts, aimed at reducing the likelihood of negative expectations, nocebo effects, data and privacy concerns, and overall social worries, regulations, and barriers to taking mHealth products to the market.

Initially, we propose that public perceptions surrounding the safety and security of mHealth technologies be systematically evaluated. Specifically, such assessments should evaluate the aforementioned biopsychosocial constructs, and recruitment should be targeted at the population of interest for mHealth implementation. For example, populations with widespread uptake of mHealth technologies (eg, high-income nations like the United States [87]) may be of value for exploring individual and societal health concerns as they relate to innovative mHealth products. Such findings may translate to informative precedent for assessing such research questions in additional developing country populations where significant funding and infrastructure improvements have been recently proposed for mHealth expansion [88].

Next, after evaluating the target populations' concerns with new mHealth technologies, we propose identifying any concerns that appear unfounded or overstated and developing evidence-based communication strategies that aim to minimize these misperceptions regarding biophysiological and psychological/behavioral risks related to mHealth implementation. In turn, these communications should enhance acceptability and self-efficacy for utilization of such technologies. Our view is that for meaningful impact, such interventions should focus on addressing the target population's most prevalent technological misperceptions and should be developed using theory-based procedures. The Health Belief Model [89], the Theory of Planned Behavior [90], and the Communication Persuasion Model [91] may be helpful evidence-based theoretical approaches to employ in this context.

Finally, we assert that evidence-informed dissemination efforts of these theory-based communications (to the population of interest) are helpful to achieve the population-level effects for *preventing* additional future mHealth concerns. Providing consumers with accurate information, knowledge, and efficacy affirmations that appropriately (ie, evidence-based) allay potential worries, enhance trust, and facilitate the overall uptake of beneficial mHealth technologies are needed for this technology reaching its potential to improve health. Such public outreach and intervention endeavors will require determining the most appropriate and wide-reaching dissemination platform for the population of interest (eg, social media vs physician communication), the optimal frequency of communications, and how to design and tailor the specific content in such communications.

## Conclusions

New technologies can afford opportunities for improving societal function and individual well-being. mHealth technology holds the potential to have a large positive global impact on health. However, scientifically unsupported concerns surrounding electromagnetic radiation exposure may produce negative health outcomes via nocebo effects; coupled with data

privacy issues and larger social implications, the array of related worries may hinder the development and implementation of mHealth technologies. We argue it is thus essential to carefully attend to these potential concerns, including distinguishing between unfounded mHealth-related worries and risks from evidence-based concerns and then evaluate and address the unfounded mHealth concerns. Some apprehension about new digital platforms (eg, wireless-powered pacemakers) may be expected, especially when the technology is very new and not widely documented or disseminated; it may also be the case that concerns, and avoidance may even serve to be proactively protective in some situations (eg, data security of PHI). Notwithstanding that some worries are entirely appropriate and adaptive, our goal here is to outline areas where unwarranted negative expectations and public concerns regarding mHealth are more likely. We hope that such attention may result in increased uptake of safe and effective mHealth, facilitating the reach and positive impact of these exciting new technologies.

This work was supported in part by the National Science Foundation Nanoscience Engineering Research Center for Advanced Self-Powered Systems of Integrated Sensors and Technologies (EEC-1160483).

## Conflicts of Interest

None declared.

## References

1. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
2. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research,Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: a scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
3. Andreu-Perez J, Leff DR, Ip HM, Yang G. From wearable sensors to smart implants--toward pervasive and personalized healthcare. IEEE Trans Biomed Eng 2015 Dec;62(12):2750-2762. [doi: 10.1109/TBME.2015.2422751] [Medline: 25879838]
4. McCallum C, Rooksby J, Gray CM. Evaluating the impact of physical activity apps and wearables: interdisciplinary review. JMIR Mhealth Uhealth 2018 Mar 23;6(3):e58 [FREE Full text] [doi: 10.2196/mhealth.9054] [Medline: 29572200]
5. Adibi S, Mobasher A, Tofigh T. LTE networking: extending the reach for sensors in mHealth applications. Trans Emerg Telecommun Technol 2014;25(7):692-706. [doi: 10.1002/ett.2598]
6. 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]
7. Baig MM, GholamHosseini H, Connolly MJ. Mobile healthcare applications: system design review, critical issues and challenges. Australas Phys Eng Sci Med 2015 Mar;38(1):23-38. [doi: 10.1007/s13246-014-0315-4] [Medline: 25476753]
8. Charani E, Gharbi M, Moore L, Castro-Sanchéz E, Lawson W, Gilchrist M, et al. Effect of adding a mobile health intervention to a multimodal antimicrobial stewardship programme across three teaching hospitals: an interrupted time series study. J Antimicrob Chemother 2017 Jun 1;72(6):1825-1831 [FREE Full text] [doi: 10.1093/jac/dkx040] [Medline: 28333297]
9. Krohn R. mHealth: a pathway to the intelligent hospital. Mhealth 2015;1:16 [FREE Full text] [doi: 10.3978/j.issn.2306-9740.2015.08.01] [Medline: 28293574]
10. Lin JC. Wireless power transfer for mobile applications, and health effects [Telecommunications Health and Safety]. IEEE Antennas Propag Mag 2013 Apr;55(2):250-253. [doi: 10.1109/MAP.2013.6529362]
11. Ghorai K, Akter S, Khatun F, Ray P. mHealth for smoking cessation programs: a systematic review. J Pers Med 2014 Jul 18;4(3):412-423 [FREE Full text] [doi: 10.3390/jpm4030412] [Medline: 25563359]
12. Kemp CG, Velloza J. Implementation of eHealth interventions across the HIV care cascade: a review of recent research. Curr HIV/AIDS Rep 2018 Dec;15(6):403-413 [FREE Full text] [doi: 10.1007/s11904-018-0415-y] [Medline: 30171519]
13. Vervloet M, Linn AJ, van Weert JC, de Bakker DH, Bouvy ML, van Dijk L. The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: a systematic review of the literature. J Am Med Inform Assoc 2012;19(5):696-704 [FREE Full text] [doi: 10.1136/amiajnl-2011-000748] [Medline: 22534082]
14. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
15. Hull TD, Mahan K. A study of asynchronous mobile-enabled SMS text psychotherapy. Telemed J E Health 2017 Mar;23(3):240-247. [doi: 10.1089/tmj.2016.0114] [Medline: 27797646]
16. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol 2015 Sep 16;3:32 [FREE Full text] [doi: 10.1186/s40359-015-0089-9] [Medline: 26376626]

17. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (integrate, design, assess, and share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
18. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
19. Materia F, Downs DS, Smyth J. Conceptualizing Acceptability in mHealth Research. In: Proceedings of the 6th Biannual Meeting of the Society for Ambulatory Assessment. 2019 Presented at: 6th Biannual Meeting of the Society for Ambulatory Assessment; June 19-22, 2019; Syracuse, NY URL: https://www.ssri.psu.edu/event/2927/6th-biannual-meeting-society-ambulatory-assessment
20. Brent RL. X-ray, microwave, and ultrasound: the real and unreal hazards. Pediatr Ann 1980 Dec;9(12):469-473. [Medline: 7005845]
21. Bren S. Historical introduction to EMF health effects. IEEE Eng Med Biol Mag 1996;15(4):24-30. [doi: 10.1109/51.511979]
22. Oyewopo AO, Olaniyi SK, Oyewopo CI, Jimoh AT. Radiofrequency electromagnetic radiation from cell phone causes defective testicular function in male Wistar rats. Andrologia 2017 Dec;49(10). [doi: 10.1111/and.12772] [Medline: 28261838]
23. Chapman S, Crichton F. Wind Turbine Syndrome: A Communicated Disease. First Edition. Sydney: Sydney University Press; 2017.
24. The Daily Mail. 2014 Mar 9. 'I Used to Be Sick All the Time': Dozens of Americans Who Claim to Be Allergic to Electromagnetic Signals Settle in Small West Virginia Town Where WiFi is Banned URL: https://www.dailymail.co.uk/news/article-2576848/I-used-sick-time-Dozens-Americans-claim-allergic-electromagnetic-signals-settle-small-West-Virginia-town-WiFi-banned.html [accessed 2019-02-15]
25. Foster KR. Scientific American Blogs. 2019 Sep 16. 5G is Coming: How Worried Should We Be about the Health Risks? URL: https://blogs.scientificamerican.com/observations/5g-is-coming-how-worried-should-we-be-about-the-health-risks/ [accessed 2019-12-02]
26. Moskowitz JM. Scientific American Blogs. 2019 Oct 17. We Have No Reason to Believe 5G is Safe URL: https://blogs.scientificamerican.com/observations/we-have-no-reason-to-believe-5g-is-safe/ [accessed 2019-12-01]
27. Di Ciaula A. Towards 5G communication systems: are there health implications? Int J Hyg Environ Health 2018 Apr;221(3):367-375. [doi: 10.1016/j.ijheh.2018.01.011] [Medline: 29402696]
28. Faasse K, Petrie KJ. The nocebo effect: patient expectations and medication side effects. Postgrad Med J 2013 Sep;89(1055):540-546. [doi: 10.1136/postgradmedj-2012-131730] [Medline: 23842213]
29. Petrie K, Wessely S. Modern worries, new technology, and medicine. Br Med J 2002 Mar 23;324(7339):690-691 [FREE Full text] [doi: 10.1136/bmj.324.7339.690] [Medline: 11909772]
30. Schmidt DC. Digital Content Next (DCN). 2018 Aug. Google Data Collection URL: https://digitalcontentnext.org/wp-content/uploads/2018/08/DCN-Google-Data-Collection-Paper.pdf [accessed 2020-03-25]
31. Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):229 [FREE Full text] [doi: 10.5210/ojphi.v5i3.4814] [Medline: 24683442]
32. Liang H, Xue Y. Avoidance of information technology threats: a theoretical perspective. Manag Inf Syst Q 2009;33(1):71-90. [doi: 10.2307/20650279]
33. King DK, Glasgow RE, Leeman-Castillo B. Reaiming RE-AIM: using the model to plan, implement, and evaluate the effects of environmental change approaches to enhancing population health. Am J Public Health 2010 Nov;100(11):2076-2084. [doi: 10.2105/AJPH.2009.190959] [Medline: 20864705]
34. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
35. Rubin GJ, Munshi JD, Wessely S. Electromagnetic hypersensitivity: a systematic review of provocation studies. Psychosom Med 2005;67(2):224-232. [doi: 10.1097/01.psy.0000155664.13300.64] [Medline: 15784787]
36. Filipkowski KB, Smyth JM, Rutchick AM, Santuzzi AM, Adya M, Petrie KJ, et al. Do healthy people worry? Modern health worries, subjective health complaints, perceived health, and health care utilization. Int J Behav Med 2010 Sep;17(3):182-188. [doi: 10.1007/s12529-009-9058-0] [Medline: 19763842]
37. Bennett P, Calman K, Curtis S, Fischbacher-Smith D. Understanding public responses to risk: issues around policy and practice. In: Bennett P, Calman K, Curtis S, Fischbacher-Smith D, editors. Risk Communication and Public Health. Oxford: Oxford University Press; 2010:3-22.
38. Schreier N, Huss A, Röösli M. The prevalence of symptoms attributed to electromagnetic field exposure: a cross-sectional representative survey in Switzerland. Soz Praventivmed 2006;51(4):202-209. [doi: 10.1007/s00038-006-5061-2] [Medline: 17193782]

39. Australian Radiation Protection and Nuclear Safety Agency. Miranda, NSW. Australia; 2019. Analysis of EMR Health Complaints Register Data 2018–2019 URL: https://www.arpansa.gov.au/sites/default/files/emr-health-complaints-register-analysis-2018-2019.pdf [accessed 2019-11-20]
40. Hussain A, Ali S, Ahmed M, Hussain S. The anti-vaccination movement: a regression in modern medicine. Cureus 2018 Jul 3;10(7):e2919 [FREE Full text] [doi: 10.7759/cureus.2919] [Medline: 30186724]
41. Ruden D, Rasouli P, Lu X. Potential long-term consequences of fad diets on health, cancer, and longevity: lessons learned from model organism studies. Technol Cancer Res Treat 2007 Jun;6(3):247-254. [doi: 10.1177/153303460700600312] [Medline: 17535033]
42. Eldridge-Thomas B, Rubin GJ. Idiopathic environmental intolerance attributed to electromagnetic fields: a content analysis of British newspaper reports. PLoS One 2013;8(6):e65713 [FREE Full text] [doi: 10.1371/journal.pone.0065713] [Medline: 23799038]
43. Hall H. Science-Based Medicine. 2015 Sep 8. Lawsuit Alleges School Wi-Fi Harmed Child with Electromagnetic Hypersensitivity URL: https://sciencebasedmedicine.org/lawsuit-alleges-school-wi-fi-harmed-child-with-electromagnetic-hypersensitivity/ [accessed 2019-01-14]
44. Gruber MJ, Palmquist E, Nordin S. Characteristics of perceived electromagnetic hypersensitivity in the general population. Scand J Psychol 2018 Aug;59(4):422-427. [doi: 10.1111/sjop.12449] [Medline: 29741795]
45. Kwon M, Kim S, Koo J, Choi J, Kim D. EHS Subjects Do Not Perceive RF EMF Emitted From Smart Phones Better Than Non-EHS Subjects. In: Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012 Presented at: EMBC'12; August 28-September 1, 2012; San Diego, CA, USA p. 2190-2193. [doi: 10.1109/embc.2012.6346396]
46. Kwon MK, Choi JY, Kim SK, Yoo TK, Kim DW. Effects of radiation emitted by WCDMA mobile phones on electromagnetic hypersensitive subjects. Environ Health 2012 Sep 21;11:69 [FREE Full text] [doi: 10.1186/1476-069X-11-69] [Medline: 22998837]
47. Kwon MS, Koivisto M, Laine M, Hämäläinen H. Perception of the electromagnetic field emitted by a mobile phone. Bioelectromagnetics 2008 Feb;29(2):154-159. [doi: 10.1002/bem.20375] [Medline: 18027840]
48. Röösli M. Radiofrequency electromagnetic field exposure and non-specific symptoms of ill health: a systematic review. Environ Res 2008 Jun;107(2):277-287. [doi: 10.1016/j.envres.2008.02.003] [Medline: 18359015]
49. Röösli M, Hug K. Wireless communication fields and non-specific symptoms of ill health: a literature review. Wien Med Wochenschr 2011 May;161(9-10):240-250. [doi: 10.1007/s10354-011-0883-9] [Medline: 21638215]
50. Rubin GJ, Hillert L, Nieto-Hernandez R, van Rongen E, Oftedal G. Do people with idiopathic environmental intolerance attributed to electromagnetic fields display physiological effects when exposed to electromagnetic fields? A systematic review of provocation studies. Bioelectromagnetics 2011 Dec;32(8):593-609. [doi: 10.1002/bem.20690] [Medline: 21769898]
51. Eltiti S, Wallace D, Russo R, Fox E. Aggregated data from two double-blind base station provocation studies comparing individuals with idiopathic environmental intolerance with attribution to electromagnetic fields and controls. Bioelectromagnetics 2015 Feb;36(2):96-107. [doi: 10.1002/bem.21892] [Medline: 25644455]
52. Barsky AJ, Saintfort R, Rogers MP, Borus JF. Nonspecific medication side effects and the nocebo phenomenon. J Am Med Assoc 2002 Feb 6;287(5):622-627. [doi: 10.1001/jama.287.5.622] [Medline: 11829702]
53. Faasse K, Helfer SG, Barnes K, Colagiuri B, Geers AL. Experimental assessment of nocebo effects and nocebo side effects: definitions, study design, and implications for psychiatry and beyond. Front Psychiatry 2019;10:396 [FREE Full text] [doi: 10.3389/fpsyt.2019.00396] [Medline: 31258492]
54. Faasse K, Huynh A, Pearson S, Geers A, Helfer S, Colagiuri B. The influence of side effect information framing on nocebo effects. Ann Behav Med 2019 Jun 4;53(7):621-629. [doi: 10.1093/abm/kay071] [Medline: 30204841]
55. Mahr A, Golmard C, Pham E, Iordache L, Deville L, Faure P. Types, frequencies, and burden of nonspecific adverse events of drugs: analysis of randomized placebo-controlled clinical trials. Pharmacoepidemiol Drug Saf 2017 Jul;26(7):731-741. [doi: 10.1002/pds.4169] [Medline: 28176407]
56. Reeves RR, Ladner ME, Hart RH, Burke RS. Nocebo effects with antidepressant clinical drug trial placebos. Gen Hosp Psychiatry 2007;29(3):275-277. [doi: 10.1016/j.genhosppsych.2007.01.010] [Medline: 17484949]
57. Häuser W, Hansen E, Enck P. Nocebo phenomena in medicine: their relevance in everyday clinical practice. Dtsch Arztebl Int 2012 Jun;109(26):459-465 [FREE Full text] [doi: 10.3238/arztebl.2012.0459] [Medline: 22833756]
58. Petrie KJ, Faasse K, Crichton F, Grey A. How common are symptoms? Evidence from a New Zealand national telephone survey. BMJ Open 2014 Jun 12;4(6):e005374 [FREE Full text] [doi: 10.1136/bmjopen-2014-005374] [Medline: 24928596]
59. Miamura K, Miyaji Y, Ohmura R. Feasibility Study on Wireless Power Transfer for Wearable Devices. In: Proceedings of the 2017 ACM International Symposium on Wearable Computers. 2017 Presented at: ISWC'17; September 11 - 15, 2017; Hawaii, Maui p. 166-167. [doi: 10.1145/3123021.3123030]
60. Strömmer E, Kaartinen J, Pärkkä J, Ylisaukko-Oja A, Korhonen I. Application of near field communication for health monitoring in daily life. Conf Proc IEEE Eng Med Biol Soc 2006;2006:3246-3249. [doi: 10.1109/IEMBS.2006.260021] [Medline: 17945761]
61. Dearing JW. Applying diffusion of innovation theory to intervention development. Res Soc Work Pract 2009 Sep 1;19(5):503-518 [FREE Full text] [doi: 10.1177/1049731509335569] [Medline: 20976022]

62. Pew Research Center. 2018. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-04-01]
63. Popper B. The Verge. 2017 May 3. A Quarter of the World's Population Now Uses Facebook Every Month URL: https://www.theverge.com/2017/5/3/15535216/facebook-q1-first-quarter-2017-earnings [accessed 2019-01-15]
64. Kharpal A. CNBC. 2016 Nov 3. Google Android Hits Market Share Record With Nearly 9 in Every 10 Smartphones Using It URL: https://www.cnbc.com/2016/11/03/google-android-hits-market-share-record-with-nearly-9-in-every-10-smartphones-using-it.html [accessed 2019-01-15]
65. Zou Y, Mhaidli AH, Mccall A, Schaub F. 'I've Got Nothing to Lose': Consumers' Risk Perceptions and Protective Actions after the Equifax Data Breach. In: Proceedings of the Fourteenth Symposium on Usable Privacy and Security. 2018 Presented at: SOUPS'18; August 12–14, 2018; Baltimore, MD p. 197-216 URL: https://www.usenix.org/system/files/conference/soups2018/soups2018-zou.pdf
66. Bhattarai A. The Washington Post. 2018 Jan 31. Privacy Experts Alarmed as Amazon Moves Into the Health Care Industry URL: https://www.washingtonpost.com/news/business/wp/2018/01/30/amazon-already-has-huge-amounts-of-our-data-what-happens-when-you-add-healthcare-to-the-mix/ [accessed 2018-12-26]
67. Rand-Hendriksen M. LinkedIn. 2015 Jul 21. Data Insecurity and What We Can Do About It URL: https://www.linkedin.com/pulse/data-insecurity-what-we-can-do-morten-rand-hendriksen [accessed 2018-12-30]
68. Andreou A, Silva M, Benevenuto F, Goga O, Loiseau P, Mislove A. Measuring the Facebook Advertising Ecosystem. In: Proceedings of the 2019 Network and Distributed Systems Security Symposium. 2019 Presented at: NDSS'19; February 24-27, 2019; San Diego, CA, USA. [doi: 10.14722/ndss.2019.23280]
69. Harwell D. The Washington Post. 2018 Aug 7. Your Banking Data Was Once Off-Limits to Tech Companies. Now They're Racing to Get It URL: https://www.washingtonpost.com/technology/2018/08/07/your-banking-data-was-once-off-limits-tech-companies-now-theyre-racing-get-it/?utm_term=.529506115594 [accessed 2019-01-15]
70. Steinfeld N. 'I agree to the terms and conditions': (How) do users read privacy policies online? An eye-tracking experiment. Comput Human Behav 2016 Feb;55:992-1000. [doi: 10.1016/j.chb.2015.09.038]
71. Budnitz ME. The Reading Room.: Publications at Reading Room; 2000. Consumers Surfing for Sales in Cyberspace: What Constitutes Acceptance and What Legal Terms and Conditions Bind the Consumer URL: https://readingroom.law.gsu.edu/cgi/viewcontent.cgi?article=1796&context=gsulr [accessed 2019-01-15]
72. Gomes R, Lapão LV. The adoption of IT security standards in a healthcare environment. Stud Health Technol Inform 2008;136:765-770. [Medline: 18487824]
73. US Food and Drug Administration. 2017. Requirements for Health Care Facilities and Manufacturers URL: https://www.fda.gov/MedicalDevices/ProductsandMedicalProcedures/GeneralHospitalDevicesandSupplies/MedicalDeviceDataSystems/ucm251927.htm [accessed 2019-02-15]
74. Medicines & Health Products Regulatory Agency. The Government of UK. 2018 Mar. 'GXP' Data Integrity Guidance and Definitions URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf [accessed 2018-12-30]
75. Alivecor. 2019. Kardia App Terms of Service URL: https://www.alivecor.com/tos/en/ [accessed 2019-12-30]
76. Polar. 2019. Terms of Use URL: https://www.polar.com/us-en/legal/terms-of-use [accessed 2019-12-30]
77. Spire Healthcare. 2017. Patient Terms and Conditions URL: https://www.spirehealthcare.com/media/17894/spire-healthcare-patient-terms-and-conditions-web-v3-nov2017-update.pdf [accessed 2019-12-30]
78. Kazdin AE. Covert modeling, model similarity, and reduction of avoidance behavior. Behav Ther 1974 May;5(3):325-340. [doi: 10.1016/s0005-7894(74)80002-x]
79. Chen D, Zhao H. Data Security and Privacy Protection Issues in Cloud Computing. In: Proceedings of the 2012 International Conference on Computer Science and Electronics Engineering. 2012 Presented at: ICCSEE'12; March 23-25, 2012; Hangzhou, China p. 647-651. [doi: 10.1109/iccsee.2012.193]
80. Chakravorty A, Wlodarczyk T, Rong C. Privacy Preserving Data Analytics for Smart Homes. In: Proceedings of the 2013 IEEE Security and Privacy Workshops. 2013 Presented at: SPW'13; May 23-24, 2013; San Francisco, CA, USA p. 23-27. [doi: 10.1109/spw.2013.22]
81. Chaffey D. Smart Insights. 2019 Feb 12. Global Social Media Research Summary 2019 URL: https://www.smartinsights.com/social-media-marketing/social-media-strategy/new-global-social-media-research/ [accessed 2019-11-15]
82. Ng V, Kent C. The Conversation. 2018 Feb 7. Smartphone Data Tracking is More Than Creepy - Here's Why You Should Be Worried URL: http://theconversation.com/smartphone-data-tracking-is-more-than-creepy-heres-why-you-should-be-worried-91110 [accessed 2019-04-01]
83. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. Manag Inf Syst Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
84. Kulviwat S, Bruner II GC, Kumar A, Nasco SA, Clark T. Toward a unified theory of consumer acceptance technology. Psychol Mark 2007 Dec;24(12):1059-1084. [doi: 10.1002/mar.20196]
85. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: A systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]

86. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):88 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]
87. Bassett DR, Freedson PS, John D. Wearable activity trackers in clinical research and practice. Kinesiol Rev 2019 Feb;8(1):11-15. [doi: 10.1123/kr.2018-0060]
88. Pillay Y, Motsoaledi PA. Digital health in South Africa: innovating to improve health. BMJ Glob Health 2018;3(Suppl 2):e000722 [FREE Full text] [doi: 10.1136/bmjgh-2018-000722] [Medline: 29713513]
89. Ahadzadeh AS, Sharif SP, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
90. Ajzen I. Web Hosting at UMass Amherst. 2019. Behavioral Interventions Based on the Theory of Planned Behavior URL: https://people.umass.edu/aizen/pdf/tpb.intervention.pdf [accessed 2019-04-01]
91. McGuire WJ. Public communication as a strategy for inducing health-promoting behavioral change. Prev Med 1984 May;13(3):299-319. [doi: 10.1016/0091-7435(84)90086-0] [Medline: 6387698]

## Abbreviations

**EHS:** electromagnetic hypersensitivity syndrome
**EMF:** electromagnetic field
**mHealth:** mobile health
**NFC:** near-field communication
**PHI:** personal health information

*Edited by G Eysenbach; submitted 12.04.19; peer-reviewed by C McCallum, F Khatun; comments to author 03.10.19; revised version received 05.12.19; accepted 09.02.20; published 25.05.20.*

*Please cite as:*
*Materia FT, Faasse K, Smyth JM*
*Understanding and Preventing Health Concerns About Emerging Mobile Health Technologies*
*JMIR Mhealth Uhealth 2020;8(5):e14375*
*URL: http://mhealth.jmir.org/2020/5/e14375/*
*doi:10.2196/14375*
*PMID:32449688*

Review

# Smart Shirts for Monitoring Physiological Parameters: Scoping Review

Hamzeh Khundaqji[1], BSc, DPT; Wayne Hing[1], PhD; James Furness[1], PhD; Mike Climstein[2,3], PhD

[1]Faculty of Health Sciences & Medicine, Bond University, Gold Coast, Australia

[2]School of Health and Human Sciences, Southern Cross University, Bilinga, Australia

[3]Physical Activity, Lifestyle, Ageing and Wellbeing Faculty Research Group, University of Sydney, Sydney, Australia

**Corresponding Author:**
Hamzeh Khundaqji, BSc, DPT
Faculty of Health Sciences & Medicine
Bond University
2 Promethean Way
Gold Coast, 4226
Australia
Phone: 61 0431443642
Email: hamzeh.khundaqji@student.bond.edu.au

## Abstract

**Background:** The recent trends of technological innovation and widescale digitization as potential solutions to challenges in health care, sports, and emergency service operations have led to the conception of smart textile technology. In health care, these smart textile systems present the potential to aid preventative medicine and early diagnosis through continuous, noninvasive tracking of physical and mental health while promoting proactive involvement of patients in their medical management. In areas such as sports and emergency response, the potential to provide comprehensive and simultaneous physiological insights across multiple body systems is promising. However, it is currently unclear what type of evidence exists surrounding the use of smart textiles for the monitoring of physiological outcome measures across different settings.

**Objective:** This scoping review aimed to systematically survey the existing body of scientific literature surrounding smart textiles in their most prevalent form, the smart shirt, for monitoring physiological outcome measures.

**Methods:** A total of 5 electronic bibliographic databases were systematically searched (Ovid Medical Literature Analysis and Retrieval System Online, Excerpta Medica database, Scopus, Cumulative Index to Nursing and Allied Health Literature, and SPORTDiscus). Publications from the inception of the database to June 24, 2019 were reviewed. Nonindexed literature relevant to this review was also systematically searched. The results were then collated, summarized, and reported.

**Results:** Following the removal of duplicates, 7871 citations were identified. On the basis of title and abstract screening, 7632 citations were excluded, whereas 239 were retrieved and assessed for eligibility. Of these, 101 citations were included in the final analysis. Included studies were categorized into four themes: (1) prototype design, (2) validation, (3) observational, and (4) reviews. Among the 101 analyzed studies, prototype design was the most prevalent theme (50/101, 49.5%), followed by validation (29/101, 28.7%), observational studies (21/101, 20.8%), and reviews (1/101, 0.1%). Presented prototype designs ranged from those capable of monitoring one physiological metric to those capable of monitoring several simultaneously. In 29 validation studies, 16 distinct smart shirts were validated against reference technology under various conditions and work rates, including rest, submaximal exercise, and maximal exercise. The identified observational studies used smart shirts in clinical, healthy, and occupational populations for aims such as early diagnosis and stress detection. One scoping review was identified, investigating the use of smart shirts for electrocardiograph signal monitoring in cardiac patients.

**Conclusions:** Although smart shirts have been found to be valid and reliable in the monitoring of specific physiological metrics, results were variable for others, demonstrating the need for further systematic validation. Analysis of the results has also demonstrated gaps in knowledge, such as a considerable lag of validation and observational studies in comparison with prototype design and limited investigation using smart shirts in pediatric, elite sports, and emergency service populations.

***(JMIR Mhealth Uhealth 2020;8(5):e18092)*** doi:10.2196/18092

**KEYWORDS**

wearable electronic devices; biomedical technology; telemedicine; fitness trackers; sports; exercise; physiology; clinical decision making; vital signs

# Introduction

## Background

Recent years have seen a marked trend of technological innovation through widescale digitization in the areas of health care, sports, and emergency operation services [1-3]. Propelled by technological progress and motivated by improving quality of care while reducing costs, the immense volume of health data produced today holds promise for aiding in ways such as clinical decision support, disease surveillance, health management, and performance optimization [1,4,5]. With a global shift toward personalized, preventative, and evidence-based models of care, the use of noninvasive monitoring and data analysis to inform clinical practice and training program design is steadily increasing [1,2,6].

Owing to the large volume, velocity, and variety of data produced, there has been a need to adopt advanced and complex technology capable of data collection, storage, and analysis [4]. Among these innovative technologies are wearable systems for physiological metrics tracking conceived by the convergence of microelectronics, wireless communication, and analytics. Designated by the Global Observatory for eHealth as mobile health systems, these forms of technology have been recognized by the World Health Organization as an essential element of electronic health, which prioritizes the cost-effective and secure use of digital technologies in support of medical and public health practice [7,8].

Wearables such as the wrist-worn Fitbit (Fitbit Inc) have garnered incredible commercial acceptance, with revenues reaching US $347 million in the third quarter of 2019 [9,10]. However, despite demonstrating validity, reliability, and acceptability for their estimates of physiological metrics such as heart rate (HR), the use of these devices is largely targeted toward fitness enthusiasts rather than researchers or clinicians [11-13].

Over the past two decades, the increased demand for noninvasive and comfortable long-term tracking of physiological metrics among clinicians has been met through an increase in the research and development of another type of wearable technology known as the smart textile.

Smart textiles are products made up of fibers, filaments, and yarns that host several electronic components, such as sensors, read-out circuits, and embedded communication systems, powered by an integrated or external power supply. Communication systems such as Bluetooth allow for the connectivity of the textile to other intelligent devices for the visualization and analysis of the data obtained in real time. These textile systems, typically designed as electronic-embedded clothing, offer a relaxed structure capable of noninvasive tracking and simultaneous communication of physiological and biomechanical data. In health care, these intelligent textile systems have the ability to support telemedicine and promote the proactive involvement of patients in their medical management through the collection and tracking of their health and diagnostic data [14]. The potential to improve preventative medicine and early diagnosis through the continuous tracking of physical and mental health status as well as physical activity also exists. [14].

In the spheres of sports and emergency response, the ability of these systems to provide continuous physiological data in real time is considerable. The detection and subsequent use of metrics indicative of the physical performance, physiological status, and mental alertness of an athlete or emergency operator have been shown to mitigate injuries and improve performance [2,15]. Although wearable systems such as those produced by Catapult (Catapult Innovations) are currently used by sporting teams to monitor workload and impact, these systems are typically limited in the physiological metrics monitored [6,15,16]. Smart textiles, on the other hand, have the potential to provide medical personnel and performance specialists with additional comprehensive, physiological insights across multiple body systems.

However, it is currently unclear what type of evidence surrounding the use of smart textiles for physiological parameter monitoring exists. For this reason, a scoping review was conducted to systematically survey the existing body of scientific literature on smart textiles in their most prevalent form, the smart shirt, for the monitoring of physiological parameters.

## Objectives

The primary outcomes were to (1) provide a clear indication of the volume and types of scientific literature relating to smart shirts, (2) summarize the studies completed to date, and (3) identify any knowledge gaps to inform future research.

To guide the review, the following research question was formulated: what is the extent, range, and nature of the scientific literature pertaining to smart shirts for physiological monitoring?

# Methods

## Protocol and Registration

An a priori protocol was developed using the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) Extension for Scoping Reviews: Checklist and Explanation [17]. The final protocol was registered prospectively with the Open Science Framework (DOI 10.17605/OSF.IO/TNK9X) on August 8, 2019 [18].

## Eligibility Criteria

The eligibility criteria were informed by the Population-Concept-Context framework recommended by the Joanna Briggs Institute (JBI) Reviewer's Manual [19].

### Population

This scoping review did not impose any restrictions on the population. Men and women of any population or age were suitable for inclusion.

### Concept

The concept of this scoping review was the monitoring of physiological outcome measures using smart textiles in the form of smart shirts. For the purpose of this review, a smart textile was defined as an intelligent textile structure or fabric that possesses integrated sensors for the monitoring and recording of physiological parameters while worn.

### Context

All study designs were considered for this scoping review, which included published articles and reviews, conference proceedings, gray literature, and chapters in the text. Studies conducted across all settings were considered for inclusion. Studies were excluded if they focused on a singular component of the smart shirt (ie, materials, sensors, or algorithms) rather than the smart shirt as an integrated unit. Furthermore, because this review focused on physiological parameters, studies concerning smart shirts for biomechanical or activity monitoring were also excluded.

### Information Sources

To identify potentially relevant literature, a 3-step approach was utilized. First, a limited preliminary search was conducted in 2 electronic bibliographic databases relevant to the topic: Ovid Medical Literature Analysis and Retrieval System Online (MEDLINE) and Excerpta Medica database (EMBASE). The limited search was then followed by analysis of the text words contained in the titles and abstracts of the retrieved papers and of the index terms used to describe them. A second comprehensive search strategy was then developed using all identified keywords and index terms by the lead investigator (HK) in consultation with a librarian highly experienced in electronic searches. Using the final search strategy, the following bibliographic databases were searched from inception of the database to June 24, 2019: Ovid MEDLINE, EMBASE, Scopus, Cumulative Index to Nursing and Allied Health Literature, and SPORTDiscus. The search results were exported into EndNote (Clarivate Analytics), with duplicates removed. Finally, the electronic database search was supplemented by scanning the reference lists of the included studies. The Canadian Agency for Drugs and Technologies in Health (CADTH) gray literature searching tool was also used to identify any nonindexed literature of relevance to this review [20].

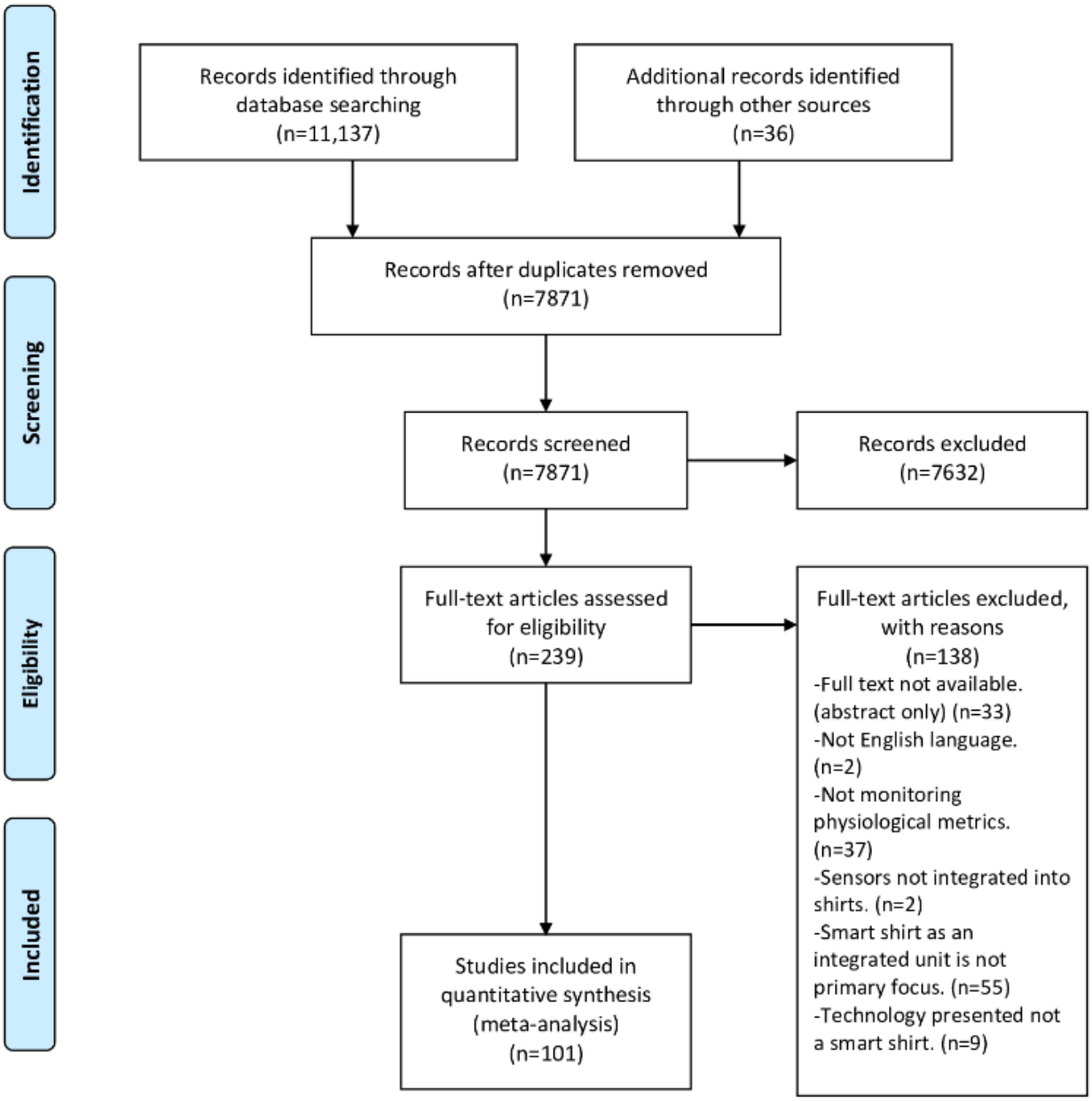
### Search

The final search strategy for all databases used can be found in Multimedia Appendix 1. Owing to the large number of irrelevant citations returned by the Scopus database, the search strategy was refined through the inclusion of additional keywords using the “AND” operator to focus the returned results on smart shirts for the monitoring of physiological outcome measures.

### Selection of Sources of Evidence

Using a priori eligibility criteria, a standardized questionnaire for study selection was developed to assist in the screening of titles, abstracts, and full text (Multimedia Appendix 2). A pilot exercise preceded each level of screening. Any queries raised by the pilot exercise were reviewed and resulted in the amendment of the questionnaire by the lead investigator. Following the removal of duplicates, the lead investigator screened papers based on title and abstract. Papers that did not meet the eligibility criteria were removed. Subsequently, the full texts of the remaining papers were retrieved and screened to determine their eligibility. As per the PRISMA guidelines, a flow diagram outlining the study selection process was produced (Figure 1). A critical appraisal of individual sources of evidence was not undertaken because this scoping review aimed to provide a map of the extent, range, and nature of the existing evidence rather than seek the best available evidence related to practice or policy [17].

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-analysis flow diagram. CINAHL: The Cumulative Index to Nursing and Allied Health; EMBASE: Excerpta Medica database; MEDLINE: Medical Literature Analysis and Retrieval System Online.




## Data Charting and Data Items

First, a data-charting form was adapted from the JBI Methodology Guidance for Scoping Reviews at the protocol stage [19] (Multimedia Appendix 3). Main areas of interest were identified, such as study citation details (eg, author, year of publication, reference type, country of origin, and study design), key study characteristics (eg, sample characteristics, study aims, type of smart shirt used, comparators, and outcomes measured), and key findings. Once the form was created, it was tested in a pilot data-charting exercise using 10 studies to ensure that all relevant data were being captured. The data extraction fields were updated through an iterative process that resulted in the inclusion of additional fields for thematic analysis (eg, types of signals acquired, and sensors used). Once the testing was complete, and the data-charting form was refined, the lead investigator (HK) independently screened all included studies and extracted key information from them. Data charted in the final extraction form included study citation details (eg, author, year of publication, reference type, country of origin, and study

design), key study characteristics (eg, sample characteristics, study aims, type of smart shirt used, types of signals acquired, types of sensors used, comparators when applicable, and outcomes measured), and key study findings.

### Synthesis of Results

Studies were categorized according to the four main themes identified: (1) prototype design, (2) validation, (3) observational, and (4) reviews. Key study characteristics and findings are graphically represented and tabulated.

## Results

### Selection of Sources of Evidence

Following the removal of duplicates, a total of 7871 citations were identified from searches of the electronic databases, the CADTH gray literature searching tool, and the reference lists of included studies. On the basis of title and abstract screening, 7632 citations were excluded, whereas 239 were retrieved and assessed for eligibility. Of these, 138 were excluded for the following reasons: 33 were abstracts or the author was unable to retrieve the full text, 2 were in Mandarin, 37 were not focused on physiological parameters collected by smart shirts, 2 included technology that failed the study's definition of a smart shirt due to the lack of integration of their sensors, 55 were focused on only one aspect of a smart shirt (eg, a sensor, materials, or algorithm) rather than the functional unit, and 9 were not smart shirts (eg, chest straps; Figure 1).

### General Study Characteristics

In sorting the included studies by publication type, journal articles were the most prevalent (60/101, 59.4%), followed by conference proceedings (37/101, 36.6%), thesis dissertations (3/101, 3.0%), and reviews (1/101, 1.0%). The years of publication identified in the literature search ranged from 1999 to 2019, with the years 2015 to 2018 producing the majority of publications (47/101, 46.5%). Publications were categorized into four themes of study: (1) prototype design, (2) validation, (3) observational, and (4) reviews. Table 1 presents the general characteristics and associated references of the analyzed publications, including year of publication, type of publication, and theme of study. The countries of origin varied widely, with 24 countries represented by 5 continents: Europe (61/101, 60.4%), North America (18/101, 17.8%), Asia (15/101, 14.9%), Australia and Oceania (4/101, 4.0%), and South America (3/101, 3.0%, Table 2). Among the 24 countries, Italy produced the bulk of the relevant literature (25/101, 24.8%).

Table 2 presents the countries represented within each continent, their respective references, and the number of studies by theme originating from each country.

**Table 1.** Characteristics of included studies (N=101).

| Characteristics | Number of studies, n | Reference(s) |
|---|---|---|
| **Year of publication** | | |
| Before 2000 | 1 | [21] |
| 2000-2004 | 2 | [22] |
| 2005-2009 | 25 | [3,23-45] |
| 2010-2014 | 18 | [46-63] |
| 2015-2018 | 47 | [64-111] |
| 2019 | 8 | [112-117] |
| **Type of publication** | | |
| Journal article | 60 | [21-23,25,28,32,34-36,39-43,45,46,49,53,56,59-65,67,69-71,73,74,76,78,80-83,85,86,88-90,93,95-97,99-101,104,106,108,109,114-119] |
| Conference proceeding | 37 | [3,24,26,27,29-31,33,37,38,44,47,48,50,54,55,57,58,68,72,75,77,79,84,87,92,94,98,102,103,105,107,110,111,113,120] |
| Thesis dissertation | 3 | [51,66,91] |
| Reviews | 1 | [112] |
| **Theme of study** | | |
| Prototype design | 50 | [3,21,22,24-26,31-34,36-40,42-44,47,50,57-60,67-71,77-80,86-90,99-105,114,118,120] |
| Validation | 29 | [27-29,41,45,46,48,52,53,61-63,72-74,81,92,93,106-111,115-117,119] |
| Observational | 21 | [23,30,35,49,54-56,64,65,75,76,82-85,94,96-98] |
| Review | 1 | [112] |

**Table 2.** Countries of origin of the included studies by total and thematic numbers (N=101).

| Continent and country of origin | Total number of studies by country, n | Number of studies by theme | | | |
|---|---|---|---|---|---|
| | | Prototype design (n=50) | Validation (n=29) | Observational (n=21) | Reviews (n=1) |
| **Europe** | | | | | |
| Belgium [32,38] | 2 | 2 | N/A[a] | N/A | N/A |
| Finland [23] | 1 | N/A | N/A | 1 | N/A |
| France [27,107] | 2 | N/A | 2 | N/A | N/A |
| Germany [42,59,60,63,81,98,105] | 7 | 4 | 2 | 1 | N/A |
| Ireland [31,47] | 2 | 2 | N/A | N/A | N/A |
| Italy [3,25,26,29,30,33,36,37,46,48,49,53,54,56,58,64,76,79,84,92,99,108,110,116,120] | 25 | 10 | 8 | 7 | N/A |
| Poland [95,100] | 2 | N/A | N/A | 2 | N/A |
| Portugal [68,71,75,80,102,113] | 6 | 6 | N/A | N/A | N/A |
| Slovakia [57] | 1 | 1 | N/A | N/A | N/A |
| Spain [39,50,62,101] | 4 | 3 | 1 | N/A | N/A |
| Switzerland [28,34,55,72,83,100] | 6 | 2 | 2 | 2 | N/A |
| United Kingdom [45,52,82] | 3 | N/A | 2 | 1 | N/A |
| **North America** | | | | | |
| Canada [35,74,87,104,114] | 5 | 3 | 1 | 1 | N/A |
| United States [21,41,61,65,66,73,85,93,96,97,106,109,117] | 13 | 1 | 8 | 4 | N/A |
| **Asia** | | | | | |
| China [40,44,78,88,90] | 5 | 5 | N/A | N/A | N/A |
| India [43] | 1 | 1 | N/A | N/A | N/A |
| Japan [70] | 1 | 1 | N/A | N/A | N/A |
| Malaysia [111] | 1 | N/A | 1 | N/A | N/A |
| South Korea [22,24,69,89,118,119] | 6 | 5 | 1 | N/A | N/A |
| Taiwan [67] | 1 | 1 | N/A | N/A | N/A |
| **Australia** | | | | | |
| Australia [112] | 1 | N/A | N/A | N/A | 1 |
| New Zealand [91,94,115] | 3 | N/A | 1 | 2 | N/A |
| **South America** | | | | | |
| Chile [77,86,103] | 3 | 3 | N/A | N/A | N/A |

[a]N/A: Not applicable.

## Study Themes

### *Prototype Design Studies*

Throughout the 50 analyzed prototype design studies, the capabilities of the presented smart shirts varied from acquiring one physiological signal (cardiac, respiratory, or surface electromyography [sEMG]) to numerous signals simultaneously (Table 3).

The physiological sensors integrated into the presented prototypes also varied considerably. Across the 50 studies, 10 distinct cardiac and respiratory sensors were identified (Table 4). Among the cardiac sensors, the most prevalent was the textile electrode, which was used in 28 different studies. The piezoresistive sensor, or strain gauge, was the most common respiratory sensor, with its use reported in 10 studies (10/50, 20%). sEMG electrodes were used in all studies (4/50, 8%), presenting a prototype capable of measuring electrical muscle activity. Other physiological sensors integrated into the prototypes included those capable of measuring body temperature (BT) and blood oxygen saturation ($SpO_2$). Table 4 presents all the different sensors identified across all prototype design studies.

**Table 3.** Types of signals acquired by the prototypes presented in the included studies (N=50).

| Signals acquired | Value, n (%) |
|---|---|
| Cardiac only [32,57,60,67-69,78,80,88,105,118] | 13 (26) |
| Respiratory only [31,42,47,51,79,87,104] | 8 (16) |
| Electromyography only [38,103] | 2 (4) |
| Numerous signals [3,22,24-26,33,34,36,37,39,40,44,50,58,59,77,80,86,99,102,113,114,120] | 27 (54) |

**Table 4.** Types and prevalence of the physiological sensors used in prototype studies (categories not exclusive; N=81)

| Classification and type of sensor | Number of studies, n | Reference(s) |
|---|---|---|
| **Cardiac** | | |
| Adhesive button electrodes | 1 | [22] |
| Bluetooth heart rate monitor | 1 | [67] |
| Noncontact, metal capacitive electrodes | 1 | [69] |
| Conductive ink electrodes | 1 | [70] |
| Disposable electrodes | 1 | [24] |
| Phonocardiogarphy | 1 | [44] |
| Photoplethysmography | 2 | [43,44] |
| Pulse sensor | 2 | [100] |
| Silicon electrodes | 2 | [43,114] |
| Textile electrodes | 28 | [3,25,26,32-34,36,37,40,50,57,59,60,68,71,77,78,80,86,88-90,99,102,105,113,118,120] |
| Not specified | 3 | [21,39,58] |
| **Respiratory** | | |
| Antenna (fiber, spiral, and hybrid spiral) | 2 | [51,57,104] |
| Fiber Bragg grating sensor | 1 | [79] |
| Impedance pneumography | 2 | [22,33] |
| Noncontact, metal capacitive electrodes | 1 | [101] |
| Optical fiber | 1 | [42] |
| Piezoresistive | 10 | [3,25,26,33,36,47,77,86,120] |
| Polypyrrole | 1 | [31] |
| Respiratory inductive plethysmography | 5 | [34,37,40,44,114] |
| Sensor coil | 1 | [59] |
| Textile | 2 | [99,102] |
| Not specified | 1 | [58] |
| **Electromyography** | | |
| Surface electromyography electrodes | 4 | [38,102,103,113] |
| **Oxygen saturation** | | |
| Pulse oximeter | 2 | [46,48] |
| Not specified | 1 | [120] |
| **Body temperature** | | |
| Monolithic | 1 | [3] |
| Bandgap | 1 | [50] |
| Digital sensor and thermistor | 1 | [43] |
| Not specified | 2 | [120] |

### *Validation Studies*

Following prototype design, validation was the most recurrent study theme. All identified studies investigated the validity of a smart shirt against an established reference technology to measure one or more physiological outcome measures. A total of 16 types of smart shirts were validated (Table 5). Among these were 3 commercially available shirts, Hexoskin (Carrè Technologies Inc), LifeShirt (LifeShirt VivoMetric), and Zephyr BioHarness (Zephyr Technology), and 13 working prototypes. Included in these prototypes was the Protection e-Textiles system developed as part of a European initiative focused on developing a wearable textile system for emergency operators [46,48]. The physiological sensors integrated into each shirt varied and ranged from combinations of cardiac, respiratory, BT, and $SpO_2$ sensors. Table 5 presents the outcome measures validated in each smart shirt.

Table 6 summarizes the physiological sensors integrated in each smart shirt. Multimedia Appendix 4 summarizes the validation studies across citation characteristics, study participants, type of smart shirt used, physiological outcome measures validated, reference technology used as a comparator, and main findings.

**Table 5.** Physiological outcome measures validated by a smart shirt across all validation studies.

| Type of smart shirt | Respiratory rate | Minute ventilation | Tidal volume | Breath duration | Heart rate | Electrocardiograph signals[a] | Body temperature | Blood oxygen saturation | Energy expenditure |
|---|---|---|---|---|---|---|---|---|---|
| BioShirt | N/A[b] | N/A | N/A | N/A | N/A | X[c] | N/A | N/A | N/A |
| GOW system | N/A | N/A | N/A | N/A | N/A | X | N/A | N/A | N/A |
| HeartCycle's guided exercise system | X | N/A | N/A | N/A | X | X | N/A | N/A | N/A |
| Hexoskin | X | X | X | N/A | X | X | N/A | N/A | X |
| Long Term Medical Survey System | X | N/A | N/A | N/A | X | N/A | X | N/A | N/A |
| Maglietta Interattiva Computerizzata | X | N/A | N/A | N/A | X | X | N/A | N/A | N/A |
| Prototype 1 [81] | N/A | N/A | N/A | N/A | X | N/A | N/A | N/A | N/A |
| Prototype 2 [52] | X | N/A | X | X | N/A | N/A | N/A | N/A | N/A |
| Prototype 3 [92] | X | N/A | X | N/A | N/A | N/A | N/A | N/A | N/A |
| Prototype 4 [108] | X | N/A | X | X | N/A | N/A | N/A | N/A | N/A |
| Prototype 5 [110] | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Protection e-Textiles (inner garment) | N/A | N/A | N/A | N/A | N/A | N/A | X | X | N/A |
| LifeShirt | X | N/A | N/A | N/A | N/A | X | N/A | N/A | N/A |
| Wealthy system | X | X | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Wearable Wellness System | N/A | N/A | N/A | N/A | N/A | X | N/A | N/A | N/A |
| Zephyr BioHarnes | N/A | N/A | N/A | N/A | X | N/A | N/A | N/A | N/A |

[a]P and T waves, QRS complex, respiratory rate intervals, and heart rate variability.

[b]N/A: not applicable.

[c]X indicates physiological outcome measures validated by studies for the corresponding smart shirt.

**Table 6.** Types of physiological sensors used by smart shirts in the identified validation studies.

| Smart shirt and category of sensor | Physiological sensors | Number of studies, n | Reference(s) |
|---|---|---|---|
| **BioShirt** | | | |
| C[a] | 1-lead ECG[b] | 1 | [119] |
| **GOW** | | | |
| C | 1-lead ECG | 1 | [62] |
| **HeartCycle's guided exercise** | | | |
| C | 1-lead ECG | 1 | [63] |
| R[c] | Not specified | 1 | [63] |
| **Hexoskin** | | | |
| C | 1-lead ECG | 10 | [65,73,74,93,106,107,109,111,115,117] |
| R | RIP[d] | 10 | [65,73,74,93,106,107,109,111,115,117] |
| **LifeShirt** | | | |
| C | 2-lead ECG | 3 | [28,41,45] |
| R | RIP | 3 | [28,41,45] |
| **Long Term Medical Survey System** | | | |
| C | 2-lead ECG | 1 | [72] |
| R | Transthoracic bioimpedance | 1 | [72] |
| $SpO_2$[e] | 4-channel optical sensor | 1 | [72] |
| BT[f] | BT | 1 | [72] |
| **Maglietta Interattiva Computerizzata** | | | |
| C | 1-lead ECG | 2 | [29,53] |
| R | Piezoresistive plethysmography | 2 | [29,53] |
| **Protection e-Textiles** | | | |
| C | 1-lead ECG | 2 | [46,48] |
| R | Piezoresistive plethysmography | 2 | [46,48] |
| $SpO_2$ | Pulse oximeter (finger) | 2 | [46,48] |
| BT | N/A[g] | 2 | [46,48] |
| **Prototype 1** | | | |
| C | 12-lead ECG | 1 | [81] |
| **Prototype 2** | | | |
| C | 1-lead ECG | 1 | [52] |
| **Prototype 3** | | | |
| R | FBG[h] | 1 | [92] |
| **Prototype 4** | | | |
| R | FBG | 1 | [108] |
| **Prototype 5** | | | |
| R | FBG | 1 | [110] |
| **Wealthy System** | | | |
| C | 5-lead ECG | 1 | [27] |
| R | Impedance pneumography | 1 | [27] |

| Smart shirt and category of sensor | Physiological sensors | Number of studies, n | Reference(s) |
|---|---|---|---|
| **Wearable Wellness System** | | | |
| C | 1-lead ECG | 1 | [116] |
| **Zephyr BioHarness** | | | |
| R | 1-lead ECG | 1 | [61] |

[a]C: cardiac.

[b]ECG: electrocardiograph.

[c]R: respiratory.

[d]RIP: respiratory inductance plethysmography.

[e]$SpO_2$: blood oxygen saturation.

[f]BT: body temperature.

[g]N/A: not applicable.

[h]FBG: fiber Bragg grating.

### *Observational Studies*

Observational studies made up approximately 21.0% (21/101) of the included publications. These studies used smart shirts to capture various physiological outcome measures in a range of populations. In total, 10 types of smart shirts were used, with Hexoskin being the most prevalent (8/21, 38%), followed by the Personalized Monitoring System for Care in Mental Health system (3/21, 14%). The experimental settings varied among the analyzed studies, with the majority (13/21, 62%) being conducted in the field. A total of 5 studies were conducted in a controlled setting such as a laboratory, whereas 2 used both controlled and free-living settings. The populations studied also varied, with the predominant population being clinical (9/21, 43%). The remaining studies used occupational populations such as medical personnel, office employees, firefighters (5/21, 24%), healthy participants (4/21, 19%), and a combination of healthy and clinical participants (3/21, 14%). All studies recruited adults with the exception of one that recruited both adults and pediatric participants [95]. Multimedia Appendix 5 summarizes the observational studies across citation characteristics, type of shirt used, study aim, population characteristics, study setting, and physiological outcome measures tracked.

### Reviews

Only one review focusing on the use of smart shirts for the monitoring of physiological parameters was identified through a literature search. This review was in the form of a scoping review with the objectives of exploring, organizing, and presenting the existing literature on the use of electronic textiles for electrocardiograph (ECG) monitoring in cardiac populations. The review identified resting ECG as the most common form of ECG acquired by electronic textiles, followed by exercise ECG and ambulatory ECG. The primary technical issue reported across all studies was noise from motion artifacts [112].

## Discussion

### Principal Findings

The primary purpose of this scoping review was to systematically analyze the body of scientific literature pertaining to smart shirts for the monitoring of physiological parameters. The primary outcomes were to (1) provide a clear indication of the volume and types of scientific literature relating to smart shirts, (2) summarize the studies completed to date, and (3) identify any knowledge gaps to inform future research.

From the 7871 citations identified after the removal of duplicates, 239 (3.00%) were eligible for full-text review. Of these 239 citations, 101 (1.3%) were included in the final study. Although the percentage of included studies appears small, this was an expected outcome due to the broad search strategy employed. The reason behind the broad search strategy was the lack of standardized terminology in the field of wearables due to its relatively recent inception. For example, smart textiles may be referred to as electronic textiles, e-textiles, electronic devices, wearable devices, wearable monitoring devices, wearable systems, etc. This required the inclusion of a wide range of keywords in the search strategy to maximize the capturing of relevant literature. However, the limitation of this strategy is that the search resulted in many studies that failed the inclusion criteria.

Throughout the screening and review process, four main themes of study were identified: (1) prototype design, (2) validation, (3) observational, and (4) reviews. The most prominent theme was prototype design, accounting for approximately 49.5% (50/101) of the total included studies. These studies presented the design of wearable systems in the form of sensor-integrated shirts for continuous and noninvasive monitoring of cardiorespiratory parameters. Although some prototypes were capable of only monitoring a single parameter (23/50, 46%), the majority (27/50, 54%) could monitor several simultaneously. These physiological parameters were classified as cardiac, respiratory, sEMG, BT, and $SpO_2$. Aside from the monitoring capabilities, it was evident that the key focus of design was on the *wearability* of the smart shirts over longer periods. Many of the established technologies used today for the monitoring of physiological parameters such as the ECG or Holter monitor are only capable of short-term diagnostic recording because of their restricted portability and uncomfortable sensors. To circumvent these issues, many of the included studies integrated their smart shirts with textile sensors. Textile sensors are intended to be comfortable, lightweight, flexible, stretchable, conformable, washable, and long lasting. This contrasts with

sensors such as the standard single-use, disposable silver-silver chloride (Ag/AgCl) electrodes, which, in combination with conductive gels, can provoke cutaneous reactions after prolonged skin contact [121]. The Ag/AgCl electrodes are also prone to damage after repeated use owing to mechanical stress [121].

Following prototype design, validation was the most identified theme. Validation studies accounted for approximately one-third (29/101, 28.7%) of the total included studies. These studies demonstrated that smart shirts were largely valid in determining cardiorespiratory parameters such as HR and respiratory rate (RR), but showed variable validity when measuring parameters such as energy expenditure (EE), minute ventilation ($V_E$), and tidal volume ($V_T$) [28,41,107,115,117]. Notably, although maximal oxygen consumption was assessed for reliability in one study, its validity was not evaluated [115]. At the current stage, smart shirts can be considered mostly valid for the measurement of certain physiological parameters under conditions of rest and submaximal activities with variable results at maximal work rate [62,63,74,93,107,115,117]. When comparing the number of prototype design and validation studies, it is evident that a large discrepancy exists. The volume of systematic validation research needs to be increased to meet the rate of prototype design studies being published. This is of importance as these wearables are marketed for use by clinicians and researchers who require valid and reliable measures. Moreover, owing to the fluid state of the software used in these devices, which can receive periodic updates, it is important to validate the new algorithm.

The third theme of study identified by this review was observational research, which comprised 20.8% (21/101) of the total studies. These studies employed a variety of smart shirts in a range of populations, including healthy and clinical as well as specific population subsets such as medical personnel, employees, and emergency operators. A total of 10 distinct smart shirts were utilized, with Hexoskin being the most prevalent. These studies used the data collected by smart shirts for various purposes, such as quantifying stress through heart rate variability (HRV) data in medical personnel and identifying early physiological markers that preceded self-injurious behavior in individuals with intellectual disabilities [35,82,96]. More observational research such as these is required to explore the implementation of smart shirts in live settings and make practical use of the data collected through its processing and analysis. This is the next obvious step in the progression of wearables research.

Finally, only one scoping review was identified in this study, focusing solely on the use of smart textiles for physiological monitoring. As presented in the Results section, this review was restricted to the use of smart textiles for ECG monitoring in cardiac patients.

### Gaps in the Literature

As alluded to in the Principal Findings section, there is a considerable lag between the publication of prototype design and validation studies. Within the area of prototype design, there remains a crucial need to develop sensing technology to further optimize sensor positioning, increase detection sensitivity, and improve the signal-to-noise ratio to assist in moving wearables past the initial prototype stage. In addition, the refinement of algorithms is needed to reduce the risk of overestimating or underestimating values. Furthermore, the diversification of the populations used in validating the prototypes is also needed. Currently, the majority of validation studies include healthy or clinical adults, whereas studies focusing on pediatric populations and specific population subsets such elite athletes and emergency operators are lacking.

Observational research is also largely underrepresented. To understand the capabilities of smart shirts outside of the laboratory, more studies of this nature need to be conducted using a variety of population samples in various settings. These studies should investigate the use of the collected data for meaningful analysis as one issue challenging wearable technology is the translation of its data to create clinically actionable insights. Such insights could take the form of physiological data being collated and reported to a health practitioner on a periodic basis to ensure the optimal management of an outpatient or the implementation of warning signals sent to the patient or health practitioner when the collected data reaches a particular threshold. Although in the sporting realm, teams have begun employing data scientists to disseminate the data into usable metrics, this is largely absent in the clinical sphere. Moving forward, observational studies would be best conducted using a multidisciplinary approach whereby researchers collaborate with the clinicians or professionals expected to implement this technology outside of the laboratory.

### Limitations

This review was limited to publications written in English, which may have excluded key studies published in other languages. This review also focused solely on physiological parameters measured by smart shirts and did not report on any other parameters concerning activity or biomechanics. In addition, the screening, inclusion/exclusion, and data charting stages of this review were conducted by 1 investigator (HK), which could have reduced the likelihood that all relevant studies were identified in the review. The use of 1 investigator may have also resulted in some reviewer bias.

### Comparisons With Previous Work

To the authors' knowledge, this scoping review is the first to systematically map the scientific body of literature surrounding the use of smart textiles in the form of smart shirts for monitoring physiological parameters across all populations. Only one other review that was included in the results investigated the use of smart shirts for monitoring a single physiological parameter [112]. The inclusion criteria of the aforementioned review were limited to smart shirts capable of monitoring ECG signals in cardiac patients, whereas this scoping review included all physiological parameters and population types.

## Conclusions

With the persistent challenges confronting health systems globally and the rising health and safety demands of athletes

and emergency operators, smart textiles present themselves as a contributor to a possible solution. This scoping review systematically surveyed the existing body of scientific literature pertaining to smart textiles in the form of smart shirts for the monitoring of physiological parameters. Through this review, it was identified that the majority of studies surrounding smart textiles were dedicated to prototype design, whereas validation and observational studies lagged behind considerably. Although smart shirts have been proven to be valid and reliable in the monitoring of some physiological parameters such as HR, HRV, and RR, results were variable for other parameters such as $V_E$, $V_T$, and EE, suggesting a continued need for their systematic validation. Although innovations such as these offer vast potential, it is important to ensure their validity and reliability through careful evaluation before their widespread adoption. To unlock the potential of smart textiles, there is also a need for more observational investigation in collaboration with the professionals expected to implement the technology outside of the laboratory.

## Authors' Contributions

HK, WH, JF, and MC conceived the scoping review. HK performed the scoping review and analyzed the data. HK wrote the paper, and WH, JF, and MC revised and provided edits contributing to the final manuscript. WH, JF, and MC reviewed and provided feedback for approval of the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies used for each electronic database searched.
[DOCX File , 21 KB - mhealth_v8i5e18092_app1.docx ]

Multimedia Appendix 2
Study selection questionnaire.
[DOCX File , 13 KB - mhealth_v8i5e18092_app2.docx ]

Multimedia Appendix 3
Data extraction chart.
[DOCX File , 14 KB - mhealth_v8i5e18092_app3.docx ]

Multimedia Appendix 4
Summaries of included validation studies.
[DOCX File , 40 KB - mhealth_v8i5e18092_app4.docx ]

Multimedia Appendix 5
Summaries of included observational studies.
[DOCX File , 35 KB - mhealth_v8i5e18092_app5.docx ]

## References

1. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014;2:3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
2. Seshadri DR, Drummond C, Craker J, Rowbottom JR, Voos JE. Wearable devices for sports: new integrated technologies allow coaches, physicians, and trainers to better understand the physical demands of athletes in real time. IEEE Pulse 2017;8(1):38-43. [doi: 10.1109/MPUL.2016.2627240] [Medline: 28129141]
3. Curone D, Dudnik G, Loriga G, Luprano J, Magenes G, Paradiso R, et al. Smart garments for safety improvement of emergency/disaster operators. Conf Proc IEEE Eng Med Biol Soc 2007;2007:3962-3965. [doi: 10.1109/IEMBS.2007.4353201] [Medline: 18002867]
4. Roski J, Bo-Linn GW, Andrews TA. Creating value in health care through big data: opportunities and policy implications. Health Aff (Millwood) 2014 Jul;33(7):1115-1122. [doi: 10.1377/hlthaff.2014.0147] [Medline: 25006136]
5. Chang H. Book review: data-driven healthcare & analytics in a big data world. Healthc Inform Res 2015;21(1):61-62 PMCID: PMC4330202. [doi: 10.4258/hir.2015.21.1.61]
6. Seshadri DR, Li RT, Voos JE, Rowbottom JR, Alfes CM, Zorman CA, et al. Wearable sensors for monitoring the internal and external workload of the athlete. NPJ Digit Med 2019;2:71 [FREE Full text] [doi: 10.1038/s41746-019-0149-2] [Medline: 31372506]

7. World Health Organization. Geneva, Switzerland: WHO; 2018. Seventy-First World Health Assembly (WHA71.7) URL: https://www.who.int/ehealth/about/en/ [accessed 2019-12-02]
8. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies: Second Global Survey on eHealth. Geneva, Switzerland: WHO Press; 2011.
9. FitBit Inc. 2019 Jun 11. Fitbit Reports Third Quarter Results for the Three Months Ended September 28, 2019 URL: https://investor.fitbit.com/press/press-releases/press-release-details/2019/Fitbit-Reports-Third-Quarter-Results-for-the-Three-Months-Ended-September-28-2019/default.aspx [accessed 2020-03-30]
10. Pantelopoulos A, Bourbakis N. A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans Syst Man Cybern C Appl Rev 2010 Jan;40(1):1-12. [doi: 10.1109/TSMCC.2009.2032660]
11. 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]
12. Thiebaud RS, Funk MD, Patton JC, Massey BL, Shay TE, Schmidt MG, et al. Validity of wrist-worn consumer products to measure heart rate and energy expenditure. Digit Health 2018;4:2055207618770322 [FREE Full text] [doi: 10.1177/2055207618770322] [Medline: 29942628]
13. Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. BMC Geriatr 2018 May 3;18(1):103 [FREE Full text] [doi: 10.1186/s12877-018-0793-4] [Medline: 29724191]
14. Wang Y, Kung L, Byrd TA. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. Technol Forecast Soc Change 2018 Jan;126:3-13. [doi: 10.1016/j.techfore.2015.12.019]
15. Li RT, Kling SR, Salata MJ, Cupp SA, Sheehan J, Voos JE. Wearable performance devices in sports medicine. Sports Health 2016;8(1):74-78 [FREE Full text] [doi: 10.1177/1941738115616917] [Medline: 26733594]
16. Wisbey B, Montgomery PG, Pyne DB, Rattray B. Quantifying movement demands of AFL football using GPS tracking. J Sci Med Sport 2010 Sep;13(5):531-536. [doi: 10.1016/j.jsams.2009.09.002] [Medline: 19897414]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
18. Khundaqji H, Hing W, Furness J, Climstein M. The Open Science Framework. 2019. Smart Shirts for Physiological Monitoring: A Scoping Review URL: https://osf.io/tnk9x/ [accessed 2019-08-19]
19. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Joanna Briggs Institute Reviewer's Manual. 2020. Chapter 11: Scoping Reviews URL: https://reviewersmanual.joannabriggs.org/ [accessed 2019-08-19]
20. Canadian Agency for Drugs and Technologies in Health. 2019 May 8. Grey Matters: A Practical Tool for Searching Health-Related Grey Literature URL: https://www.cadth.ca/resources/finding-evidence/grey-matters [accessed 2019-06-01]
21. Gopalsamy C, Park S, Rajamanickam R, Jayaraman S. The wearable motherboard: the first generation of adaptive and responsive textile structures (ARTS) for medical applications. Virtual Real 1999 Sep;4(3):152-168. [doi: 10.1007/bf01418152]
22. Shin S, Ryu CY, Kang JH, Nam SH, Song YS, Lim TG, et al. Realization of an e-health system to perceive emergency situations. Conf Proc IEEE Eng Med Biol Soc 2004;2004:3309-3312. [doi: 10.1109/IEMBS.2004.1403930] [Medline: 17270989]
23. Halín N, Junnila M, Loula P, Aarnio P. The LifeShirt system for wireless patient monitoring in the operating room. J Telemed Telecare 2005;11(Suppl 2):S41-S43. [doi: 10.1258/135763305775124623] [Medline: 16375793]
24. Nam S, Yim T, Ryu C, Shin S, Kang J, Kim S. The preliminary study of unobtrusive respiratory monitoring for e-health. Conf Proc IEEE Eng Med Biol Soc 2005;2005:3796-3798. [doi: 10.1109/IEMBS.2005.1617311] [Medline: 17281056]
25. Paradiso R, Loriga G, Taccini N. A wearable health care system based on knitted integrated sensors. IEEE Trans Inf Technol Biomed 2005 Sep;9(3):337-344. [doi: 10.1109/titb.2005.854512] [Medline: 16167687]
26. Paradiso R, Loriga G, Taccini N. Wearable system for vital signs monitoring. Stud Health Technol Inform 2004;108:253-259. [doi: 10.1109/ITAB.2003.1222533] [Medline: 15718653]
27. Bourdon L, Coli S, Loriga G, Taccini N, Gros B, Gemignani A, et al. First Results With the Wealthy Garment Electrocardiogram Monitoring System. In: Proceedings of the 2005 Computers in Cardiology Conference. 2005 Presented at: CIC'05; September 25-28, 2005; Lyon, France. [doi: 10.1109/cic.2005.1588176]
28. Clarenbach CF, Senn O, Brack T, Kohler M, Bloch KE. Monitoring of ventilation during exercise by a portable respiratory inductive plethysmograph. Chest 2005 Sep;128(3):1282-1290. [doi: 10.1378/chest.128.3.1282] [Medline: 16162719]
29. di Rienzo M, Rizzo F, Parati G, Ferratini M, Brambilla G, Castiglioni P. A Textile-Based Wearable System for Vital Sign Monitoring: Applicability in Cardiac Patients. In: Proceedings of the 2005 Computers in Cardiology Conference. 2005 Presented at: CIC'05; September 25-28, 2005; Lyon, France. [doi: 10.1109/cic.2005.1588199]
30. di Rienzo M, Rizzo F, Meriggi P, Bordoni B, Brambilla G, Ferratini M, et al. Applications of a textile-based wearable system for vital signs monitoring. Conf Proc IEEE Eng Med Biol Soc 2006;2006:2223-2226. [doi: 10.1109/IEMBS.2006.260173] [Medline: 17946505]

31. Brady S, Carson B, O'Gorman D, Moyna N, Diamond D. Combining Wireless With Wearable Technology for the Development of On-Body Networks. In: Proceedings of the International Workshop on Wearable and Implantable Body Sensor Networks. 2006 Presented at: BSN'06; April 3-5, 2006; Cambridge, MA, USA. [doi: 10.1109/BSN.2006.16]
32. Coosemans J, Hermans B, Puers R. Integrating Wireless ECG Monitoring in Textiles. In: Proceedings of the 13th International Conference on Solid-State Sensors, Actuators and Microsystems. 2005 Presented at: SENSOR'05; June 5-9, 2005; Seoul, South Korea. [doi: 10.1109/SENSOR.2005.1496399]
33. Paradiso R, de Rossi D. Advances in textile technologies for unobtrusive monitoring of vital parameters and movements. Conf Proc IEEE Eng Med Biol Soc 2006;2006:392-395. [doi: 10.1109/IEMBS.2006.259307] [Medline: 17946827]
34. Wilhelm FH, Pfaltz MC, Grossman P. Continuous electronic data capture of physiology, behavior and experience in real life: towards ecological momentary assessment of emotion. Interact Comput 2006 Mar;18(2):171-186. [doi: 10.1016/j.intcom.2005.07.001]
35. Barrera FJ, Violo RA, Graver EE. On the form and function of severe self-injurious behavior. Behav Intervent 2007 Feb;22(1):5-33. [doi: 10.1002/bin.228]
36. Castiglioni P, Faini A, Parati G, di Rienzo M. Wearable seismocardiography. Conf Proc IEEE Eng Med Biol Soc 2007;2007:3954-3957. [doi: 10.1109/IEMBS.2007.4353199] [Medline: 18002865]
37. di Rienzo M, Parati G, Rizzo F, Meriggi P, Merati G, Faini A, et al. Heart rate monitoring and control in altered gravity conditions. Conf Proc IEEE Eng Med Biol Soc 2007;2007:6682-6685. [doi: 10.1109/IEMBS.2007.4353893] [Medline: 18003559]
38. Taelman J, Adriaensen T, van der Horst C, Linz T, Spaepen A. Textile integrated contactless EMG sensing for stress analysis. Conf Proc IEEE Eng Med Biol Soc 2007;2007:3966-3969. [doi: 10.1109/IEMBS.2007.4353202] [Medline: 18002868]
39. Villalba E, Arredondo M, Ottaviano M, Salvi D, Hoyo-Barbolla E, Guillen S. Heart failure monitoring system based on wearable and information technologies. J Commun 2007 Mar 1;2(2):10-21. [doi: 10.4304/jcm.2.2.10-21]
40. Zheng JW, Zhang ZB, Wu TH, Zhang Y. A wearable mobihealth care system supporting real-time diagnosis and alarm. Med Biol Eng Comput 2007 Sep;45(9):877-885. [doi: 10.1007/s11517-007-0221-y] [Medline: 17619091]
41. Heilman KJ, Porges SW. Accuracy of the LifeShirt (Vivometrics) in the detection of cardiac rhythms. Biol Psychol 2007 Jul;75(3):300-305 [FREE Full text] [doi: 10.1016/j.biopsycho.2007.04.001] [Medline: 17540493]
42. D'Angelo L, Weber S, Honda Y, Thiel T, Narbonneau F, Luth TC. A system for respiratory motion detection using optical fibers embedded into textiles. Conf Proc IEEE Eng Med Biol Soc 2008;2008:3694-3697. [doi: 10.1109/IEMBS.2008.4650011] [Medline: 19163514]
43. Pandian PS, Mohanavelu K, Safeer KP, Kotresh TM, Shakunthala DT, Gopal P, et al. Smart Vest: wearable multi-parameter remote physiological monitoring system. Med Eng Phys 2008 May;30(4):466-477. [doi: 10.1016/j.medengphy.2007.05.014] [Medline: 17869159]
44. Zhang Z, Wang W, Wang B, Wu H, Liu H, Zhang Y. A Prototype of Wearable Respiration Biofeedback Platform and Its Preliminary Evaluation on Cardiovascular Variability. In: Proceedings of the 3rd International Conference on Bioinformatics and Biomedical Engineering. 2009 Presented at: ICBBE'09; June 11-13, 2009; Beijing, China. [doi: 10.1109/icbbe.2009.5162230]
45. Kent L, O'Neill B, Davison G, Nevill A, Elborn JS, Bradley JM. Validity and reliability of cardiorespiratory measurements recorded by the LifeShirt during exercise tests. Respir Physiol Neurobiol 2009 Jun 30;167(2):162-167. [doi: 10.1016/j.resp.2009.03.013] [Medline: 19505672]
46. Curone D, Secco EL, Tognetti A, Loriga G, Dudnik G, Risatti M, et al. Smart garments for emergency operators: the ProeTEX project. IEEE Trans Inf Technol Biomed 2010 May;14(3):694-701. [doi: 10.1109/TITB.2010.2045003] [Medline: 20371413]
47. Carlos R, Coyle S, Corcoran B, Diamond D, Tomas W, Aaron M, et al. Web-Based Sensor Streaming Wearable for Respiratory Monitoring Applications. In: Proceedings of the 2011 SENSORS Conference. 2011 Presented at: ICSENS'11; October 28-31, 2011; Limerick, Ireland. [doi: 10.1109/icsens.2011.6127168]
48. Magenes G, Curone D, Secco EL, Bonfiglio A. Biosensing and environmental sensing for emergency and protection e-textiles. Conf Proc IEEE Eng Med Biol Soc 2011;2011:8365-8368. [doi: 10.1109/IEMBS.2011.6092063] [Medline: 22256287]
49. Migliorini M, Mendez MO, Bianchi AM. Study of heart rate variability in bipolar disorder: linear and non-linear parameters during sleep. Front Neuroeng 2011;4:22 [FREE Full text] [doi: 10.3389/fneng.2011.00022] [Medline: 22291638]
50. Talavera G, Martin R, Rodríguez-Alsina A, Garcia J, Fernández F, Carrabina J. Protecting Firefighters with Wearable Devices. In: Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence. 2012 Presented at: UCAmI'12; December 3-5, 2012; Vitoria-Gasteiz, Spain p. 470-477. [doi: 10.1007/978-3-642-35377-2_65]
51. Ratnarathorn S. Semantic Scholar. Gothenburg, Sweden: Chalmers University of Technology; 2012. Design of a Stress Monitor Based on Breathing Signals Using a Smart Textile Shirt URL: https://pdfs.semanticscholar.org/a2bd/aa838d5333e6d674297d4bf71b0fd709fed1.pdf [accessed 2020-04-03]
52. Cleland I, Nugent C, Finlay D, Burns W, Bougourd J, Armitage R. Assessment of Custom Fitted Heart Rate Sensing Garments whilst undertaking Everyday Activities. In: Proceedings of the International Conference on Smart Homes and

Health Telematics. 2012 Presented at: ICOST'12; June 12-15, 2012; Artiminio, Italy p. 124-131. [doi: 10.1007/978-3-642-30779-9_16]
53. di Rienzo M, Racca V, Rizzo F, Bordoni B, Parati G, Castiglioni P, et al. Evaluation of a textile-based wearable system for the electrocardiogram monitoring in cardiac patients. Europace 2013 Apr;15(4):607-612. [doi: 10.1093/europace/eus368] [Medline: 23258818]
54. di Rienzo M, Vaini E, Castiglioni P, Lombardi P, Meriggi P, Rizzo F. A textile-based wearable system for the prolonged assessment of cardiac mechanics in daily life. Conf Proc IEEE Eng Med Biol Soc 2014;2014:6896-6898. [doi: 10.1109/EMBC.2014.6945213] [Medline: 25571581]
55. Mirmohamadsadeghi L, Fallet S, Buttu A, Saugy J, Saugy T, Heinzer R, et al. Sleep Apnea Detection Using Features From the Respiration and the ECG Recorded With Smart-Shirts. In: Proceedings of the 2014 Biomedical Circuits and Systems Conference. 2014 Presented at: BioCAS'14; October 22-24, 2014; Lausanne, Switzerland. [doi: 10.1109/biocas.2014.6981645]
56. Valenza G, Nardelli M, Lanatà A, Gentili C, Bertschy G, Paradiso R, et al. Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis. IEEE J Biomed Health Inform 2014 Sep;18(5):1625-1635. [doi: 10.1109/JBHI.2013.2290382] [Medline: 24240031]
57. Hanic M, Sládek L, Horínek F, Jagelka M, Donoval M, Dařícek M, et al. BIO-Monitoring System With Conductive Textile Electrodes Integrated Into T-Shirt. In: Proceedings of the 24th International Conference Radioelektronika. 2014 Presented at: Radioelek'14; April 15-16, 2014; Bratislava, Slovakia. [doi: 10.1109/radioelek.2014.6828472]
58. Sardini E, Serpelloni M. T-Shirt for Vital Parameter Monitoring. In: Proceedings of the First National Conference on Sensors. 2014 Presented at: SENSORS'14; February 15-17, 2014; Rome p. 201-205. [doi: 10.1007/978-1-4614-3860-1_35]
59. Teichmann D, Kuhn A, Leonhardt S, Walter M. The MAIN shirt: a textile-integrated magnetic induction sensor array. Sensors (Basel) 2014 Jan 9;14(1):1039-1056 [FREE Full text] [doi: 10.3390/s140101039] [Medline: 24412900]
60. Ulbrich M, Mühlsteff J, Sipilä A, Kamppi M, Koskela A, Myry M, et al. The IMPACT shirt: textile integrated and portable impedance cardiography. Physiol Meas 2014 Jun;35(6):1181-1196. [doi: 10.1088/0967-3334/35/6/1181] [Medline: 24846072]
61. Dolezal BA, Boland DM, Carney J, Abrazado M, Smith DL, Cooper CB. Validation of heart rate derived from a physiological status monitor-embedded compression shirt against criterion ECG. J Occup Environ Hyg 2014;11(12):833-839. [doi: 10.1080/15459624.2014.925114] [Medline: 24896644]
62. Romagnoli M, Alis R, Guillen J, Basterra J, Villacastin JP, Guillen S. A novel device based on smart textile to control heart's activity during exercise. Australas Phys Eng Sci Med 2014 Jun;37(2):377-384. [doi: 10.1007/s13246-014-0271-z] [Medline: 24756693]
63. Skobel E, Martinez-Romero A, Scheibe B, Schauerte P, Marx N, Luprano J, et al. Evaluation of a newly designed shirt-based ECG and breathing sensor for home-based training as part of cardiac rehabilitation for coronary artery disease. Eur J Prev Cardiol 2014 Nov;21(11):1332-1340. [doi: 10.1177/2047487313493227] [Medline: 23733743]
64. Lanata A, Valenza G, Nardelli M, Gentili C, Scilingo EP. Complexity index from a personalized wearable monitoring system for assessing remission in mental health. IEEE J Biomed Health Inform 2015 Jan;19(1):132-139. [doi: 10.1109/JBHI.2014.2360711] [Medline: 25291802]
65. Montes J, Stone TM, Manning JW, McCune D, Tacad DK, Young JC, et al. Using hexoskin wearable technology to obtain body metrics during trail hiking. Int J Exerc Sci 2015;8(4):425-430 [FREE Full text] [doi: 10.1249/01.mss.0000477034.13469.88] [Medline: 27293503]
66. Montes J. Digital Scholarship. 2015. Validation and Reliability of the Hexoskin and FitBit Flex Wearable BIO Collection Devices URL: https://digitalscholarship.unlv.edu/thesesdissertations/2393/ [accessed 2020-04-03]
67. Hsiao C, Lee R, Tien S, Feng Y, Huang S. Early clinical prognosis for high-risk chest pain patients using smart textiles. Biomed Eng Appl Basis Commun 2015;27(6):1550057. [doi: 10.4015/s101623721550057x]
68. Lage JE, Catarino AP, Carvalho H, Rocha A. Smart Shirt With Embedded Vital Sign and Moisture Sensing. In: Proceedings of the First International Conference on Smart Portable, Wearable, Implantable and Disability-oriented Devices and Systems. 2015 Presented at: SPWID'15; June 21 - 26, 2015; Brussels, Belgium URL: https://www.semanticscholar.org/paper/Smart-shirt-with-embedded-vital-sign-and-moisture-Lage-Catarino/e22968b24987b6bce82f9cd67b5918a93885b896
69. Mathias DN, Kim S, Park J, Joung Y. Real time ECG monitoring through a wearable smart t-shirt. Trans Electr Electron Mater 2015 Feb 25;16(1):16-19. [doi: 10.4313/teem.2015.16.1.16]
70. Tada Y, Amano Y, Sato T, Saito S, Inoue M. A smart shirt made with conductive ink and conductive foam for the measurement of electrocardiogram signals with unipolar precordial leads. Fibers 2015 Nov 3;3(4):463-477. [doi: 10.3390/fib3040463]
71. Trindade I, Martins F, Dias R, Oliveira C, da Silva JM. Novel textile systems for the continuous monitoring of vital signals: design and characterization. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:3743-3746. [doi: 10.1109/EMBC.2015.7319207] [Medline: 26737107]
72. Chételat O, Ferrario D, Proença M, Porchet JA, Falhi A, Grossenbacher O, et al. Clinical validation of LTMS-S: a wearable system for vital signs monitoring. Conf Proc IEEE Eng Med Biol Soc 2015;2015:3125-3128. [doi: 10.1109/EMBC.2015.7319054] [Medline: 26736954]

73. Tanner EA, Montes J, Manning JW, Taylor JE, DeBeliso M, Young JC, et al. Validation of Hexoskin biometric shirt to COSMED K4 b2 metabolic unit in adults during trail running. Sport Technol 2015;8(3-4):118-123. [doi: 10.1080/19346182.2016.1248973]
74. Villar R, Beltrame T, Hughson RL. Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. Appl Physiol Nutr Metab 2015 Oct;40(10):1019-1024. [doi: 10.1139/apnm-2015-0140] [Medline: 26360814]
75. Paiva J, Rodrigues S, Cunha JP. Changes in ST, QT and RR ECG intervals during acute stress in firefighters: a pilot study. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:3378-3381. [doi: 10.1109/EMBC.2016.7591452] [Medline: 28269028]
76. Valenza G, Nardelli M, Lanata' A, Gentili C, Bertschy G, Kosel M, et al. Predicting mood changes in bipolar disorder through heartbeat nonlinear dynamics. IEEE J Biomed Health Inform 2016 Jul;20(4):1034-1043. [doi: 10.1109/JBHI.2016.2554546] [Medline: 28113920]
77. Aqueveque P, Gutierrez C, Saavedra F, Pino EJ. Noninvasive health condition monitoring device for workers at high altitudes conditions. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:2349-2352. [doi: 10.1109/EMBC.2016.7591201] [Medline: 28268797]
78. Dai M, Xiao X, Chen X, Lin H, Wu W, Chen S. A low-power and miniaturized electrocardiograph data collection system with smart textile electrodes for monitoring of cardiac function. Australas Phys Eng Sci Med 2016 Dec;39(4):1029-1040. [doi: 10.1007/s13246-016-0483-5] [Medline: 27743384]
79. Massaroni C, Ciocchetti M, di Tomaso G, Saccomandi P, Caponero MA, Polimadei A, et al. Design and preliminary assessment of a smart textile for respiratory monitoring based on an array of Fiber Bragg Gratings. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:6054-6057. [doi: 10.1109/EMBC.2016.7592109] [Medline: 28269633]
80. Trindade IG, da Silva JM, Miguel R, Pereira M, Lucas J, Oliveira L, et al. Design and evaluation of novel textile wearable systems for the surveillance of vital signals. Sensors (Basel) 2016 Sep 24;16(10):1573 [FREE Full text] [doi: 10.3390/s16101573] [Medline: 27669263]
81. Boehm A, Yu X, Neu W, Leonhardt S, Teichmann D. A novel 12-lead ECG t-shirt with active electrodes. Electron 2016 Nov 8;5(4):75. [doi: 10.3390/electronics5040075]
82. Chatty L, Omurtag A, Roy R, Garbey M. Tracking Mental Workload by Multimodal Measurements During Minimally Invasive Surgery Training. In: Proceedings of the 2017 Scientific Session of the Society of American Gastrointestinal and Endoscopic Surgeons. 2017 Presented at: SAGES'17; March 22-25, 2017; Houston, Texas URL: https://www.sages.org/meetings/annual-meeting/abstracts-archive/tracking-mental-workload-by-multimodal-measurements-during-minimally-invasive-surgery-training/ [doi: 10.1007/s00464-017-5445-9]
83. Palix J, Akselrod M, Cungi C, Giuliani F, Favrod J. Changes in heart rate variability recorded in natural situation with t-shirt integrated sensors and level of observed behavioral excitation: a pilot study of patients with intellectual disabilities and psychiatric disorders. Front Psychiatry 2017;8:4 [FREE Full text] [doi: 10.3389/fpsyt.2017.00004] [Medline: 28203208]
84. Presti DL, Massaroni C, Formica D, Giurazza F, Schena E, Saccomandi P, et al. Respiratory and Cardiac Rates Monitoring During MR Examination by a Sensorized Smart Textile. In: Proceedings of the International Instrumentation and Measurement Technology Conference. 2017 Presented at: I2MTC'17; May 22-25, 2017; Turin, Italy. [doi: 10.1109/i2mtc.2017.7969818]
85. Webster E, Sukaviriya N, Chang H, Kozloski J. Predicting cognitive states from wearable recordings of autonomic function. IBM J Res Dev 2017 Mar 1;61(2/3):2:1-2:11. [doi: 10.1147/jrd.2017.2648698]
86. Aqueveque P, Gutierrez C, Rodriguez FS, Pino EJ, Morales AS, Wiechmann EP. Monitoring physiological variables of mining workers at high altitude. IEEE Trans Ind Appl 2017 May;53(3):2628-2634. [doi: 10.1109/tia.2017.2675360]
87. Gorgutsa S, Bellemare-Rousseau S, Guay P, Miled A, Messaddeq Y. Smart T-Shirt With Wireless Respiration Sensor. In: Proceedings of the 2017 Conference on IEEE Sensors. 2017 Presented at: ICSENS'17; October 29-November 1, 2017; Glasgow, UK. [doi: 10.1109/icsens.2017.8234293]
88. Zhang H, Tian L, Lu H, Zhou M, Zou H, Fang P, et al. A wearable 12-lead ECG acquisition system with fabric electrodes. Conf Proc IEEE Eng Med Biol Soc 2017 Jul;2017:4439-4442. [doi: 10.1109/EMBC.2017.8037841] [Medline: 29060882]
89. Park J, Han H, Heo J, Lee J. Computer aided diagnosis sensor integrated outdoor shirts for real time heart disease monitoring. Comput Assist Surg (Abingdon) 2017 Dec;22(Supl 1):176-185. [doi: 10.1080/24699322.2017.1389396] [Medline: 29037055]
90. Sun F, Yi C, Li W, Li Y. A wearable H-shirt for exercise ECG monitoring and individual lactate threshold computing. Comput Ind 2017 Nov;92-93:1-11. [doi: 10.1016/j.compind.2017.06.004]
91. Baig MM. AUT Library. Auckland, New Zealand: Auckland University of Technology; 2017. Early Detection and Self-Management of Long-Term Conditions Using Wearable Technologies URL: https://openrepository.aut.ac.nz/handle/10292/10999?show=full [accessed 2020-04-03]
92. Lo Presti D, Massaroni C, Saccomandi P, Caponero MA, Formica D, Schena E. A wearable textile for respiratory monitoring: feasibility assessment and analysis of sensors position on system response. Conf Proc IEEE Eng Med Biol Soc 2017 Jul;2017:4423-4426. [doi: 10.1109/EMBC.2017.8037837] [Medline: 29060878]
93. Montoye AH, Mitrzyk JR, Molesky MJ. Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. Meas Phys Educ Exerc Sci 2017 Jun 8;21(4):201-211. [doi: 10.1080/1091367x.2017.1331166]
94. Baig M, Mirza F, GholamHosseini H, Gutierrez J, Ullah E. Clinical decision support for early detection of prediabetes and type 2 diabetes mellitus using wearable technology. Conf Proc IEEE Eng Med Biol Soc 2018 Jul;2018:4456-4459. [doi: 10.1109/EMBC.2018.8513343] [Medline: 30441340]

95. Balsam P, Lodziński P, Tymińska A, Ozierański K, Januszkiewicz L, Główczyńska R, et al. Study design and rationale for biomedical shirt-based electrocardiography monitoring in relevant clinical situations: ECG-shirt study. Cardiol J 2018;25(1):52-59 [FREE Full text] [doi: 10.5603/CJ.a2017.0102] [Medline: 28840587]
96. Slamon NB, Penfil SH, Nadkarni VM, Parker RM. A prospective pilot study of the biometrics of critical care practitioners during live patient care using a wearable 'smart shirt'. J Intensive Crit Care 2018;4(2):10. [doi: 10.21767/2471-8505.100112]
97. Thompson JF, Severson RL, Rosecrance JC. Occupational physical activity in brewery and office workers. J Occup Environ Hyg 2018 Sep;15(9):686-699. [doi: 10.1080/15459624.2018.1492136] [Medline: 30188781]
98. Yu X, Boehm A, Neu W, Venema B, Marx N, Leonhardt S, et al. A Wearable 12-Lead ECG T-Shirt With Textile Electrodes for Unobtrusive Long-Term Monitoring–Evaluation of an Ongoing Clinical Trial. In: Proceedings of the 2017 European Medical and Biological Engineering Confernce. 2017 Presented at: EMBEC'17; June 11-15, 2017; Tampere, Finland p. 703-706. [doi: 10.1007/978-981-10-5122-7_176]
99. di Rienzo M, Vaini E, Lombardi P. Development of a smart garment for the assessment of cardiac mechanical performance and other vital signs during sleep in microgravity. Sens Actuators A Phys 2018 May;274:19-27. [doi: 10.1016/j.sna.2018.02.034]
100. Kańtoch E. Recognition of sedentary behavior by machine learning analysis of wearable sensors during activities of daily living for telemedical assessment of cardiovascular risk. Sensors (Basel) 2018 Sep 24;18(10):E3219 [FREE Full text] [doi: 10.3390/s18103219] [Medline: 30249987]
101. Naranjo-Hernández D, Talaminos-Barroso A, Reina-Tosina J, Roa L, Barbarov-Rostan G, Cejudo-Ramos P, et al. Smart vest for respiratory rate monitoring of COPD patients based on non-contact capacitive sensing. Sensors (Basel) 2018 Jul 3;18(7):2144 [FREE Full text] [doi: 10.3390/s18072144] [Medline: 29970861]
102. Paiva A, Ferreira F, Catarino A, Carvalho M, Carvalho H. Design of smart garments for sports and rehabilitation. IOP Conf Ser Mater Sci Eng 2018 Dec 7;459:12083. [doi: 10.1088/1757-899X/459/1/012083]
103. Pino E, Arias Y, Aqueveque P. Wearable EMG shirt for upper limb training. Conf Proc IEEE Eng Med Biol Soc 2018 Jul;2018:4406-4409. [doi: 10.1109/EMBC.2018.8513107] [Medline: 30441329]
104. Roudjane M, Bellemare-Rousseau S, Khalil M, Gorgutsa S, Miled A, Messaddeq Y. A portable wireless communication platform based on a multi-material fiber sensor for real-time breath detection. Sensors (Basel) 2018 Mar 25;18(4):973 [FREE Full text] [doi: 10.3390/s18040973] [Medline: 29587396]
105. Walzer T, Thies C, Meier K, Madrid NM. Textile Sensor Platform (TSP)-Development of a Textile Real-Time Electrocardiogram. In: Proceedings of the 2018 International Conference on Bioinformatics and Biomedical Engineering. 2018 Presented at: IWBBIO'18; April 25-27, 2018; Granada, Spain p. 359-370. [doi: 10.1007/978-3-319-78759-6_33]
106. Banerjee T, Peterson M, Oliver Q, Froehle A, Lawhorne L. Validating a commercial device for continuous activity measurement in the older adult population for dementia management. Smart Health (Amst) 2018 Jan;5-6:51-62 [FREE Full text] [doi: 10.1016/j.smhl.2017.11.001] [Medline: 29915807]
107. Cherif NH, Mezghani N, Gaudreault N, Ouakrim Y, Mouzoune I, Boulay P. Physiological Data Validation of the Hexoskin Smart Textile. In: Proceedings of the 11th International Conference on Biomedical Electronics and Devices. 2018 Presented at: BIODEVICES'18; January 19-21, 2018; Funchal, Portugal p. 150-156. [doi: 10.5220/0006588001500156]
108. Massaroni C, Venanzi C, Silvatti AP, Lo Presti D, Saccomandi P, Formica D, et al. Smart textile for respiratory monitoring and thoraco-abdominal motion pattern evaluation. J Biophotonics 2018 May;11(5):e201700263. [doi: 10.1002/jbio.201700263] [Medline: 29297202]
109. Montes J, Young JC, Tandy R, Navalta JW. Reliability and validation of the hexoskin wearable bio-collection device during walking conditions. Int J Exerc Sci 2018;11(7):806-816 [FREE Full text] [doi: 10.1519/JSC.0000000000002005] [Medline: 30338022]
110. Presti DL, Massaroni C, Schena PS, Formica D, Caponero MA, di Tomaso G. Smart Textile Based on FBG Sensors for Breath-by-Breath Respiratory Monitoring: Tests on Women. In: Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications. 2018 Presented at: MeMeA'18; June 11-13, 2018; Rome, Italy. [doi: 10.1109/memea.2018.8438721]
111. Tan DY, Yong TH. Suitability of SmartShirt by Hexoskin to Monitor Heart Rate for Racket Sports. In: Proceedings of the 2017 International Conference on Robotics, Automation and Sciences. 2017 Presented at: ICORAS'17; November 27-29, 2017; Melaka, Malaysia. [doi: 10.1109/icoras.2017.8308057]
112. Teferra MN, Kourbelis C, Newman P, Ramos JS, Hobbs D, Clark RA, et al. Electronic textile electrocardiogram monitoring in cardiac patients: a scoping review protocol. JBI Database System Rev Implement Rep 2019 Feb;17(2):147-156. [doi: 10.11124/JBISRIR-2017-003630] [Medline: 30204712]
113. Paiva A, Catarino A, Carvalho H, Postolache O, Postolache G, Ferreira F. Design of a Long Sleeve T-Shirt with ECG and EMG for Athletes and Rehabilitation Patients. In: Proceedings of the 2018 International Conference on Innovation, Engineering and Entrepreneurship. 2018 Presented at: HELIX'18; June 27-29, 2018; Guimarães, Portugal p. 244-250. [doi: 10.1007/978-3-319-91334-6_34]
114. 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]

115. Elliot CA, Hamlin MJ, Lizamore CA. Validity and reliability of the hexoskin wearable biometric vest during maximal aerobic power testing in elite cyclists. J Strength Cond Res 2019 May;33(5):1437-1444. [doi: 10.1519/JSC.0000000000002005] [Medline: 28759538]
116. Reali P, Tacchino G, Rocco G, Cerutti S, Bianchi AM. Heart rate variability from wearables: a comparative analysis among standard ECG, a smart shirt and a wristband. Stud Health Technol Inform 2019;261:128-133. [doi: 10.3233/978-1-61499-975-1-128] [Medline: 31156103]
117. Smith CM, Chillrud SN, Jack DW, Kinney P, Yang Q, Layton AM. Laboratory validation of hexoskin biometric shirt at rest, submaximal exercise, and maximal exercise while riding a stationary bicycle. J Occup Environ Med 2019 Apr;61(4):e104-e111. [doi: 10.1097/JOM.0000000000001537] [Medline: 30946696]
118. Lee Y, Chung W. Wireless sensor network based wearable smart shirt for ubiquitous health and activity monitoring. Sens Actuators B Chem 2009 Jul;140(2):390-395. [doi: 10.1016/j.snb.2009.04.040]
119. Hong S, Yang Y, Kim S, Shin S, Lee I, Jang Y, et al. Performance study of the wearable one-lead wireless electrocardiographic monitoring system. Telemed J E Health 2009 Mar;15(2):166-175. [doi: 10.1089/tmj.2008.0071] [Medline: 19292626]
120. Magenes G, Curone D, Lanati M, Secco EL. Long-distance monitoring of physiological and environmental parameters for emergency operators. Conf Proc IEEE Eng Med Biol Soc 2009;2009:5159-5162. [doi: 10.1109/IEMBS.2009.5332737] [Medline: 19963886]
121. Ankhili A, Tao X, Cochrane C, Coulon D, Koncar V. Washable and reliable textile electrodes embedded into underwear fabric for electrocardiography (ECG) monitoring. Materials (Basel) 2018 Feb 7;11(2):256 [FREE Full text] [doi: 10.3390/ma11020256] [Medline: 29414849]

## Abbreviations

**Ag:** silver
**AgCl:** silver chloride
**BT:** body temperature
**CADTH:** Canadian Agency for Drugs and Technologies in Health
**ECG:** electrocardiograph
**EE:** energy expenditure
**EMBASE:** Excerpta Medica database
**HR:** heart rate
**HRV:** heart rate variability
**JBI:** Joanna Briggs Institute
**MEDLINE:** Medical Literature Analysis and Retrieval System Online
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-analysis
**RR:** respiratory rate
**sEMG:** surface electromyography
**$SpO_2$:** blood oxygen saturation
**VE:** minute ventilation
**VT:** tidal volume

*Edited by G Eysenbach; submitted 03.02.20; peer-reviewed by M DeBeliso, K Pirzada; comments to author 07.03.20; revised version received 10.03.20; accepted 22.03.20; published 27.05.20.*

*Please cite as:*
*Khundaqji H, Hing W, Furness J, Climstein M*
*Smart Shirts for Monitoring Physiological Parameters: Scoping Review*
*JMIR Mhealth Uhealth 2020;8(5):e18092*
*URL: http://mhealth.jmir.org/2020/5/e18092/*
*doi:10.2196/18092*
*PMID:32348279*

Original Paper

# A Mobile Technology for Collecting Patient-Reported Physical Activity and Distress Outcomes: Cross-Sectional Cohort Study

Miyeon Jung[1*], BS; SaeByul Lee[2*], MD, PhD; Jisun Kim[2], MD, PhD; HeeJeong Kim[2], MD, PhD; BeomSeok Ko[2], MD, PhD; Byung Ho Son[2], MD, PhD; Sei-Hyun Ahn[2], MD, PhD; Yu Rang Park[3], PhD; Daegon Cho[1], PhD; Haekwon Chung[4], BSc; Hye Jin Park[2], RN; Minsun Lee[2], RN; Jong Won Lee[2], MD, PhD; Seockhoon Chung[5], MD, PhD; Il Yong Chung[2], MD

[1]Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea

[2]Division of Breast Surgery, Department of Surgery, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

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

[4]Swallaby, Seoul, Republic of Korea

[5]Department of Psychiatry, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea

[*]these authors contributed equally

**Corresponding Author:**
Il Yong Chung, MD
Division of Breast Surgery
Department of Surgery
University of Ulsan College of Medicine, Asan Medical Center
88 Olympic-ro 43-gil, Songpa-gu
Seoul,
Republic of Korea
Phone: 82 2 3010 3998
Email: doorkeeper1@gmail.com

## Abstract

**Background:** Electronic patient-reported outcome (PROs) provides a fast and reliable assessment of a patient's health-related quality of life. Nevertheless, using PRO in the traditional paper format is not practical for clinical practice due to the limitations associated with data analysis and management. A questionnaire app was developed to address the need for a practical way to group and use distress and physical activity assessment tools.

**Objective:** The purpose of this study was to assess the level of agreement between electronic (mobile) and paper-and-pencil questionnaire responses.

**Methods:** We validated the app version of the distress thermometer (DT), International Physical Activity Questionnaire (IPAQ), and Patient Health Questionnaire–9 (PHQ-9). A total of 102 participants answered the paper and app versions of the DT and IPAQ, and 96 people completed the PHQ-9. The study outcomes were the correlation of the data between the paper-and-pencil and app versions.

**Results:** A total of 106 consecutive breast cancer patients were enrolled and analyzed for validation of paper and electronic (app) versions. The Spearman correlation values of paper and app surveys for patients who responded to the DT questionnaire within 7 days, within 3 days, and on the same day were .415 ($P$<.001), .437 ($P$<.001), and .603 ($P$<.001), respectively. Similarly, the paper and app survey correlation values of the IPAQ total physical activity metabolic equivalent of task (MET; Q2-6) were .291 ($P$=.003), .324 ($P$=.005), and .427 ($P$=.01), respectively. The correlation of the sum of the Patient Health Questionnaire–9 (Q1-9) according to the time interval between the paper-based questionnaire and the app-based questionnaire was .469 for 14 days ($P$<.001), .574 for 7 days ($P$<.001), .593 for 3 days ($P$<.001), and .512 for the same day ($P$=.03). These were all statistically significant. Similarly, the correlation of the PHQ (Q10) value according to the time interval between the paper-based questionnaire and the app-based questionnaire was .283 for 14 days ($P$=.005), .409 for 7 days ($P$=.001), .415 for 3 days ($P$=.009), and .736 for the same day ($P$=.001). These were all statistically significant. In the overall trend, the shorter the interval between the paper-and-pencil questionnaire and the app-based questionnaire, the higher the correlation value.

**Conclusions:** The app version of the distress and physical activity questionnaires has shown validity and a high level of association with the paper-based DT, IPAQ (Q2-6), and PHQ-9. The app-based questionnaires were not inferior to their respective paper

versions and confirm the feasibility for their use in clinical practice. The high correlation between paper and mobile app data allows the use of new mobile apps to benefit the overall health care system.

**Trial Registration:** ClinicalTrials.gov NCT03072966; https://clinicaltrials.gov/ct2/show/NCT03072966

***(JMIR Mhealth Uhealth 2020;8(5):e17320)*** doi:10.2196/17320

**KEYWORDS**

telemedicine; breast neoplasms; mobile apps; quality of life; validation; patient-reported outcome measures (PROMs); questionnaire

## Introduction

The National Comprehensive Cancer Network recommends that whenever a cancer patient visits a doctor, the doctor should screen for distress, which can be managed by clinical practice guidelines. Similarly, the American Society of Clinical Oncology guidelines suggest all that cancer patients should be screened for depressive symptoms at appropriate intervals at the beginning of and after the visit [1]. Several papers reported that the prevalence of depression and anxiety among cancer survivors was 11.6% and 17.9%, respectively [2].

Patient-reported outcome (PRO) measures, defined by the US Food and Drug Administration as "reporting on the health of patients directly from patients," is becoming common in the medical field [3]. However, conventional distress screening tools are paper-and-pencil questionnaires, which can cause recall bias and do not reflect real-time episodes of distress. In addition, the use of PRO in traditional paper format is not practical for clinical practice due to limitations associated with data analysis and management [4,5]. Therefore, entering PRO data by electronic means (ePRO) was developed as an alternative [6]. Initially, ePRO was developed based on a web platform, thus offering the portability and viability of tools used for health care assessment via mobile phones [7,8].

Current long-term health care monitoring of patients requires the most promising remote monitoring techniques to provide cost-effective quality control [9]. Therefore, models for remote monitoring of patients combined with the selection of ePRO are recommended. This enables self-management of patient care at all treatment stages while improving the quality of life of the patient [10]. The benefit of the clinical use of the mobile phone app is the possibility to accumulate high-quality and reliable data and archive backups to prevent data loss [11].

Although some discrepancies were reported between the paper-and-pencil and electronic versions of the same questionnaires, there is evidence that electronic and paper PROs reflect equivalent outcomes, whereas some reports suggest that the electronic PRO is more accurate [8,12]. Despite the active use of mobile health (mHealth) apps for measuring distress and physical activity, no validated health care apps have been developed. The questionnaire app was developed to address the need for a practical way to group and use distress and physical activity assessment tools. According to the study guidelines proposed by the International Society for Pharmacoeconomics and Outcomes Research, the data obtained from ePRO questionnaires should be comparable or superior to the data from paper-based questionnaires [13].

Therefore, this study aims to validate the app using correlation analysis between the paper-based gold standard and mobile-based new formats using distress and physical activity questionnaires. Our study examined the association between responses collected through a mobile app and the paper-based face-to-face survey. Even though the same questionnaires were used, responses collected through the new format of mobile technology can be different than responses on paper-based questionnaires. This may be because of screen size or not having the constraint of a face-to-face survey.

As a result, if the mobile app survey was completed on the same day, even though the collection method was different, the responses were almost the same. But responses weren't similar for questions like asking about time spent sitting on the International Physical Activity Questionnaire (IPAQ). In this aspect, we could conclude that estimates of time spent sitting are not easy to answer correctly and not easy to remember, so it varies depending on the format of collection and time zone, even in the same day. Therefore, IPAQ (Q7; sitting time) is difficult to replace through mobile app collection systems.

Beyond validating the mobile responses that were answered on the same day, we also compared them with the average of mobile responses in the days before and after a few days. The reason we should compare them is that the distress thermometer (DT), IPAQ, and Patient Health Questionnaire–9 (PHQ-9) are asking the status of patients for last 7 days, not on the one single day when the survey was placed.

Therefore, we analyzed how the average values of survey inputs from the mobile-based PRO collection system, answered on the 3 or 7 days before and after the paper-based face-to-face survey, were related to the input value from the paper survey. When we want to replace a face-to-face questionnaire with a mobile app in the real world, it may be more effective to use the average value at specific intervals due to the nature of PROs asking about the status of the recent week, not one day. In our analysis, the correlation became weaker as the intervals became longer, but it still showed a significant correlation.

## Methods

### Study Design and Subjects

This cross-sectional study recruited patients who underwent surgery for breast cancer at the Asan Medical Center. Patients were eligible for study participation if they were women between the ages of 20 and 65 years and had Android smartphones compatible with the WalkON app [14], a free activity tracking app modified for this study. Patients who had distant metastasis, recurrent breast cancer, severe medical conditions such as

cardiovascular disease, did not know how to use a smartphone, used iOS smartphones, or were on chemotherapy were excluded.

Written informed consent was obtained from all patient subjects at enrollment. The study protocol was approved by the institutional review board of Asan Medical Center (2016-0819). This study was registered at ClinicalTrials.gov [NCT03072966].

### Ethical Approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

### Recruitment and Follow-Up

During the hospital stay after breast cancer surgery, subjects were contacted by a clinical research assistant. After consenting to participate, participants completed the DT, IPAQ, and PHQ-9 paper-based questionnaires (baseline). The assistant helped patients download the Android-based app (WalkON) to the participants' smartphones. The main purpose of this study was to calculate the Spearman correlation of baseline values of DT, IPAQ, and PHQ-9 between paper-based questionnaire and app-based questionnaire.

At the 3- and 6-month follow-ups, participants were asked to complete the same version of the paper surveys voluntarily. Since there were few voluntary answers, we only used the first survey to validate the correlation between app and paper surveys.

### Smartphone App and App-Based Questionnaire

Swallaby Inc is a mobile health care app company that has developed a health-related smartphone app (WalkON). This app provides users with a platform for tracking their daily steps and creates mobile communities where users can communicate with each other and view each other's daily step count to get motivated and promote health-related activities.

App-based self-reporting questionnaires were programmed into this app, wherein study participants could answer the app-based questionnaires. We used this app to conduct weekly and biweekly questionnaires for a future study on the development of a distress screening tool. Daily questionnaires were developed and previously reported by the authors and consisted of self-reporting modules for daily anxiety, sleep, and emotion statuses [9,12]. Responses were collected every week for DT and IPAQ and biweekly for PHQ-9 through the app, and push notifications were sent every week from Sunday to Tuesday to subjects' smartphones. The responses to these three questionnaires (DT, IPAQ, and PHQ-9) were analyzed in this validation study.

### Statistical Analysis

The walking app-based physical activity and stress collecting systems were validated by calculating the Spearman and concordance correlation between the responses to app-based and paper-based questionnaires. The baseline paper survey was administered once at the recruitment stage. Thereafter, participants voluntarily reported their physical activity and stress through the app. For DT and IPAQ, the survey required participants to record their average stress level and physical activity level scores, respectively, during the last 7 days. For PHQ-9, the survey required participants to record their subjective perception of average life quality level during the last 14 days.

We expected the values collected via paper and mobile app to be almost the same if they were reported on the same day. To infer that the new channel for patient-reported outcome collection, a mobile app, works well, the correlation between two values reported on the same day should be high.

However, since the patients responded to DT and IPAQ at least once per week and PHQ-9 biweekly on any day, the subsample of patients who responded to the questionnaires on the same day (recruitment date) was small.

Thus, to investigate the validity of the mobile app survey, we also calculated the correlation between two values: the values at the baseline date and values reported on a few days around the baseline date. More specifically, we picked the values recorded on the baseline date ±3 days and calculated the correlation between the values reported on and those reported around the baseline date. Similarly, we calculated the correlation between the values recorded within and after 7 days from the baseline date. For the PHQ-9 only, we also calculated the correlation between the values recorded within and after 14 days from the baseline date. Here, few participants (less than 10%) reported each value more than two times within the weekly and biweekly periods. In such cases, we calculated the average value of the reported numbers within the defined period.

Taken together, for DT and IPAQ, we first calculated the average of the values reported on the same day, within 3 days, and within 7 days. Then, we calculated the Spearman correlation between the average value in the mobile app and the paper survey answer. For PHQ-9, we calculated the average of the values reported on the same day, within 3 days, within 7 days, and within 14 days.

In addition to the Spearman correlation, we also calculated the required sample size and actual statistical power using G*Power software, given the value of $\alpha$ as .05, the power (1-$\beta$) as .95, and the Spearman correlation value.

We also conducted a robustness check on our correlation analysis by calculating the bias correction factor of the concordance correlation coefficient (CCC) [15] and by positing an ordered logistic regression model. For the CCC, previous studies suggest the following descriptive scale for values (for continuous variables): <.90 poor, .90 to .95 moderate, .95 to .99 substantial, and >.99 almost perfect correlation [15]. For the ordered logistic regression model, dependent variables were the discrete variables of DT, IPAQ and PHQ in the paper survey, and independent variables were the continuous variables of average DT, IPAQ, and PHQ in the mobile app survey. We calculated the coefficients on the values of mobile app survey and their odds ratios.

## Results

### Patient Characteristics

From June 2017 to January 2018, we consecutively assessed 1247 patients who underwent breast cancer surgery for study eligibility (Figure 1). After screening, 591 patients were excluded, 176 patients could not be contacted during the hospital stay, and 320 patients refused to join the study. A total of 160 patients were enrolled in this study. Among them, 54 patients did not complete the mobile app survey within 14 days after the baseline date. Thus, we included 106 patients who responded to the app survey at least once within the defined time period. The example of app screenshots of the survey is shown in Figure 2.

**Figure 1.** Participant enrollment.

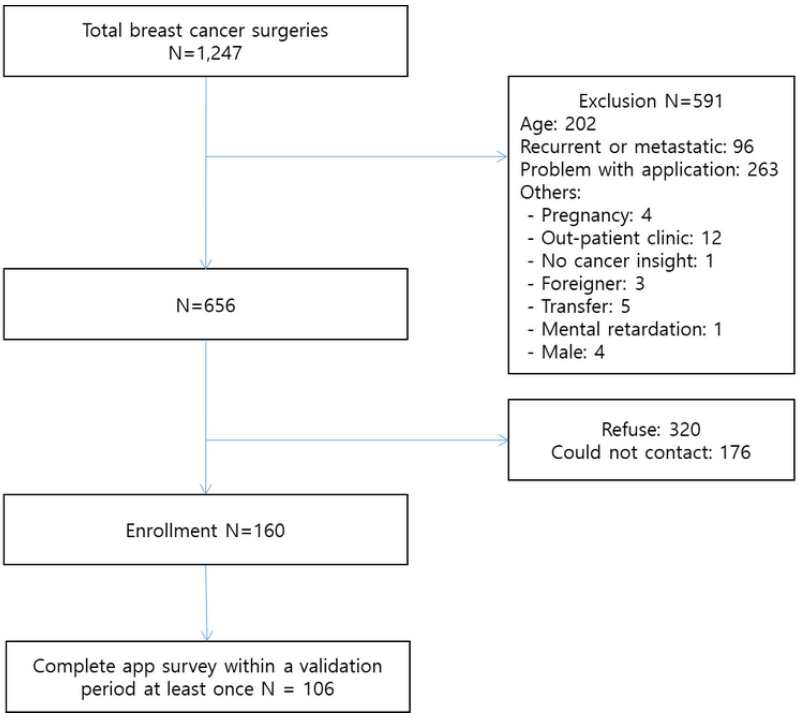


**Figure 2.** App screenshots of the (A) Distress Thermometer, (B) International Physical Activity Questionnaire, and (C) Patient Health Questionnaire–9.

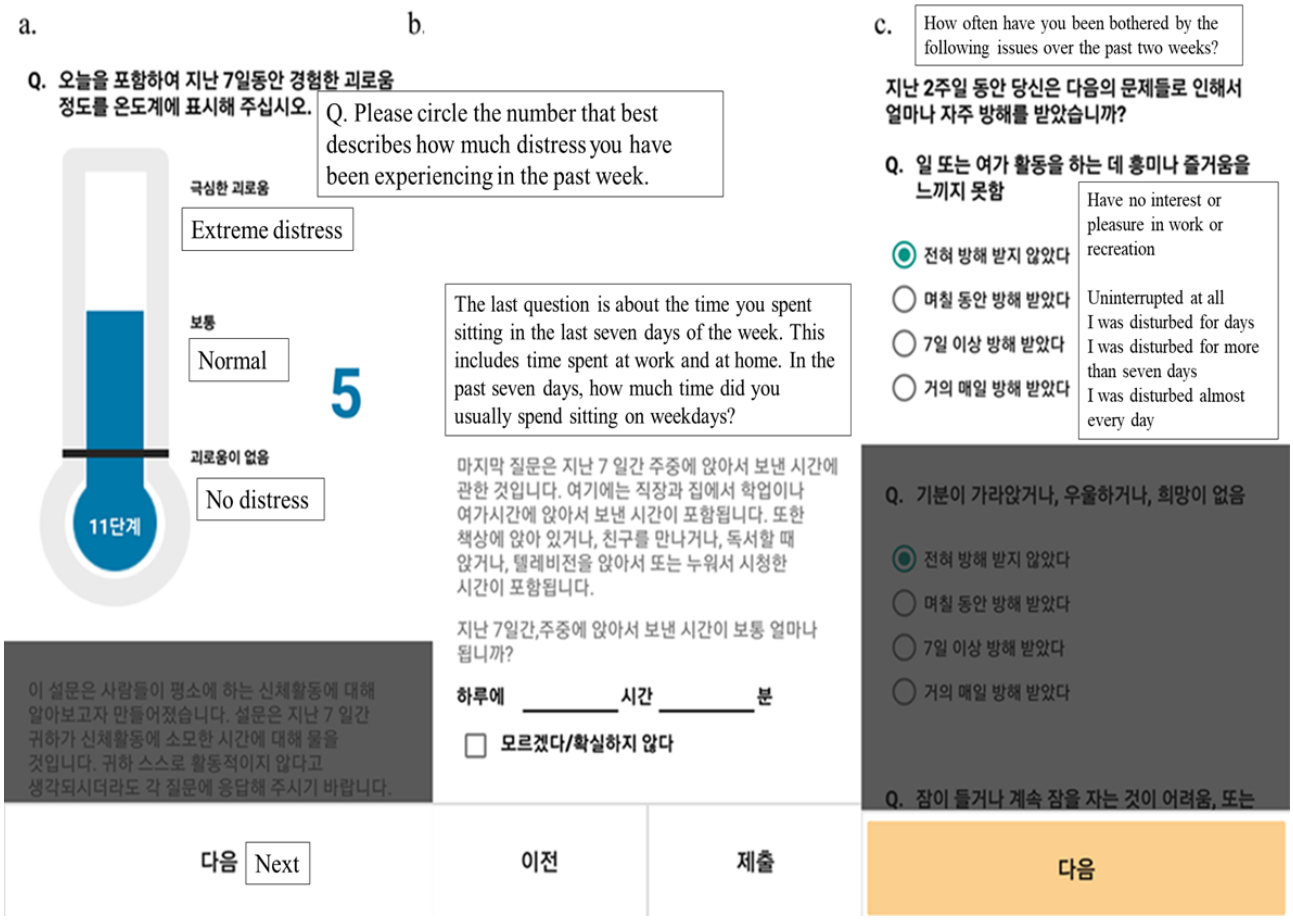



Table 1 summarizes the demographic and clinical characteristics of the subjects as absolute and relative frequencies. The subjects were aged 44.9 (SD 7.1) years. Among the patients, 77.4% (82/106) were aged less than 50 years, 64.2% (68/106) had an educational attainment of college level or higher, and 46.2% (49/106) were currently employed. Among breast cancer stages, 10.4% (11/106) of patients had stage 0, 45.3% (48/106) had stage I, 26.4% (28/106) had stage II, and 17.9% (19/106) had stage III disease (Table 1). Fifty patients completed adjuvant or neoadjuvant chemotherapy before beginning the data collection.

**Table 1.** Subject demographics (n=106).

| Characteristic | Value |
|---|---|
| **Age in years, mean (SD)** | 44.9 (7.1) |
| <50, n (%) | 82 (77.4) |
| ≥50, n (%) | 24 (22.6) |
| **Marital status, n (%)** | |
| Married | 90 (84.9) |
| Single | 14 (13.2) |
| Others | 2 (1.9) |
| **Education, n (%)** | |
| ≤High school | 38 (35.8) |
| >High school | 68 (64.2) |
| **Employed, n (%)** | |
| Yes | 49 (46.2) |
| No | 57 (53.8) |
| **Comorbidity, n (%)** | |
| Yes | 72 (67.9) |
| No | 34 (32.1) |
| **Past episode of depression, n (%)** | |
| Yes | 1 (0.9) |
| No | 101 (95.3) |
| No response | 4 (3.8) |
| **Surgery, n (%)** | |
| Mastectomy | 7 (6.6) |
| Breast-conserving surgery | 74 (69.8) |
| Mastectomy with reconstruction | 25 (23.6) |
| **Chemotherapy, n (%)** | |
| Yes | 50 (47.2) |
| No | 56 (52.8) |
| **Antihormonal therapy, n (%)** | |
| Yes | 86 (81.1) |
| No | 20 (18.9) |
| **Radiation therapy, n (%)** | |
| Yes | 86 (81.1) |
| No | 20 (18.9) |
| **Targeted therapy, n (%)** | |
| Yes | 98 (92.5) |
| No | 8 (7.5) |
| **Stage, n (%)** | |
| 0 | 11 (10.4) |
| I | 48 (45.3) |
| II | 28 (26.4) |
| III | 19 (17.9) |
| **Distress thermometer, n (%)** | |

| Characteristic | Value |
|---|---|
| Score of 5 or higher | 35 (33.0) |
| Score of less than 5 | 71 (67.0) |
| **PHQ-9[a] total score, n (%)** | |
| Score of 11 or higher | 23 (21.7) |
| Score of less than 11 | 83 (78.3) |

[a]PHQ-9: Patient Health Questionnaire–9.

## Validating the Mobile App Survey

Among 106 patients, 102 patients completed the weekly app survey (DT and IPAQ) within 7 days after the paper survey visit. Among them, 73 patients completed the weekly app survey within 3 days after the paper survey visit. In addition, 34 patients responded to the app-based questionnaire on the same day that they answered the paper-based questionnaire (Table 2). Of the 106 patients, 96 responded to the app-based questionnaire within 14 days of completing the paper-based questionnaire (PHQ-9); 63 patients answered within 7 days, 39 within 3 days, and 18 patients answered the app-based questionnaire on the same day that they answered the paper-based questionnaire.

The screening tool for measuring distress gives a numerical representation of the degree of distress (Figure 3A), IPAQ (Q2-6) asks about activity levels, and IPAQ (Q7) asks about time spent sitting (Figure 3B). IPAQ (Q2-6) and IPAQ (Q7) were analyzed separately in Table 2 because the contents of each question are different.

Spearman correlation values between the average values in app-based questionnaire being responded to within 7 days, within 3 days and on the same day and the values in the baseline paper survey were .415 ($P$<.001), .437 ($P$<.001), and .603 ($P$<.001), respectively, and the correlation values for IPAQ (Q2-6) were .291 ($P$=.003), .324 ($P$=.005), and .427 ($P$=.01), respectively. These were all statistically significant.

However, the correlation values of IPAQ (Q7) were .061 ($P$=.54) and .090 ($P$=.45) within 7 days and 3 days, respectively, and were not statistically significant. The correlation value, responded on the same day as the paper survey, was .155 ($P$=.38), which was also statistically insignificant (Table 2).

In terms of required sample size (Table 3), all required sample sizes to validate the correlation between the weekly app survey and corresponding paper survey were smaller than our sample sizes (102, 73, and 34) for DT. For IPAQ total physical activity metabolic equivalent of task (MET; Q2-6), the required sample sizes were slightly larger than our sample size. Even though our sample sizes were slightly smaller than the required sample sizes, the correlation coefficients were statistically significant.

**Table 2.** Spearman correlation coefficients between the value on the paper survey and the average value on the weekly app survey.

| Survey | Days before and after the paper survey, number of patients reporting through the app, and scores | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Same day (n=34) | | | 3 days (n=73) | | | 7 days (n=102) | | |
| | Correlation | S-statistic | *P* value | Correlation | S-statistic | *P* value | Correlation | S-statistic | *P* value |
| Distress thermometer | .603 | 2601 | <.001 | .437 | 36,511 | <.001 | .415 | 103,473 | <.001 |
| IPAQ[a] total physical activity MET[b] (Q2-6) | .427 | 3753.1 | .01 | .324 | 43,790 | .005 | .291 | 125,385 | .003 |
| IPAQ sitting MET (Q7) | .155 | 5530.3 | .38 | .090 | 58,977 | .45 | .061 | 166,062 | .54 |

[a]IPAQ: International Physical Activity Questionnaire.

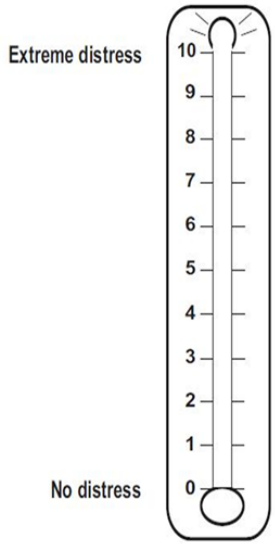
[b]MET: metabolic equivalent of task.

**Figure 3.** Paper-based versions of the (A) Distress Thermometer, (B) International Physical Activity Questionnaire, and (C) Patient Health Questionnaire–9.

a.

Screening tool for measuring distress

Instructions: First, please circle the number (0-10) that best describes how much distress you have been experiencing in the past week including today.




b.

Short form of IPAQ

1. During the last 7 days, on how many days did you do vigorous physical activities like heavy lifting, digging, aerobics, or fast bicycling?
   _____days per week
2. How much time did you usually spend doing vigorous physical activities on one of those days?
   _____hours per day _____minutes per day
3. During the last 7 days, on how many days did you do moderate physical activities like carrying light loads, bicycling at a regular pace, or doubles tennis? Do not include walking.
   _____days per week
4. How much time did you usually spend doing moderate physical activities on one of those days?
   _____hours per day _____minutes per day
5. During the last 7 days, on how many days did you walk for at least 10 minutes at a time?
   _____days per week
6. How much time did you usually spend walking on one of those days?
   _____hours per day _____minutes per day
7. During the last 7 days, how much time did you spend sitting on a week day?
   _____hours per day _____minutes per day

c.

How often have you been bothered by any of the following problems? Circle your response.

| | Not at all | Some | Often | Nearly all of the time |
|---|---|---|---|---|
| Little interest or pleasure in doing things | 0 | 1 | 2 | 3 |
| Feeling down, depressed, or hopeless | 0 | 1 | 2 | 3 |
| Trouble failing or staying asleep, or sleeping to much | 0 | 1 | 2 | 3 |
| Feeling tired or having little energy | 0 | 1 | 2 | 3 |
| Poor appetite or overeating | 0 | 1 | 2 | 3 |
| Feeling bad about yourself – or that you are a failure or have let your family down | 0 | 1 | 2 | 3 |
| Trouble concentrating on things, such as reading the newspaper or watching television | 0 | 1 | 2 | 3 |
| Moving or speaking so slowly that other people could have noticed. Or the opposite – being so fidgety or restless that you have been moving around a lot more than usual | 0 | 1 | 2 | 3 |
| Thoughts that you would be better off dead, or of hurting yourself | 0 | 1 | 2 | 3 |

**Table 3.** Sample size and actual statistical power in the correlation analysis of weekly app survey.

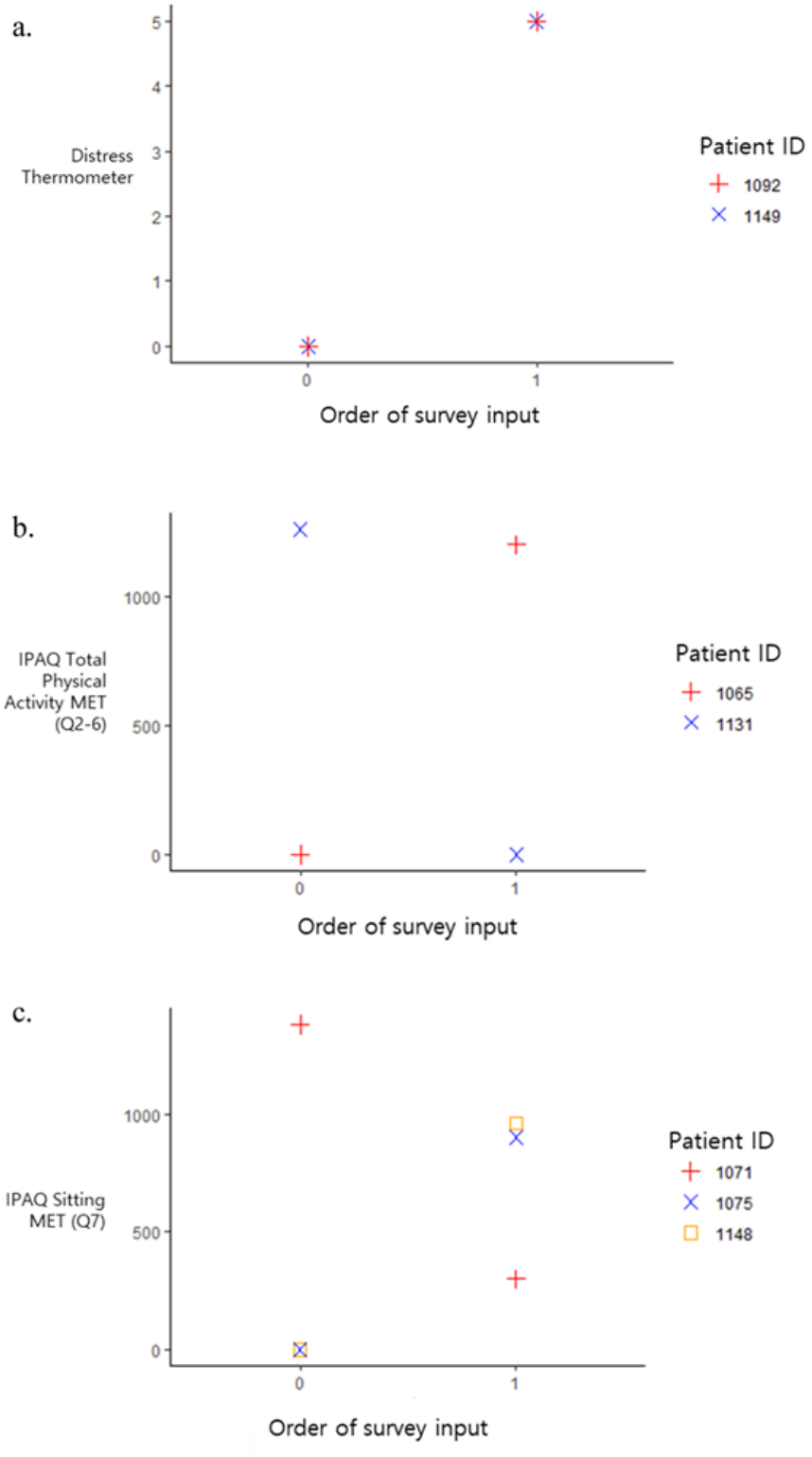
| Survey | Days before and after the paper survey and scores | | | | | |
|---|---|---|---|---|---|---|
| | Same day | | 3 days | | 7 days | |
| | Sample size | Actual power | Sample size | Actual power | Sample size | Actual power |
| Distress Thermometer | 21 | .955 | 52 | .952 | 58 | .952 |
| IPAQ[a] total physical activity MET[b] (Q2-6) | 50 | .962 | 98 | .950 | 123 | .951 |
| IPAQ sitting MET (Q7) | 446 | .950 | 1331 | .950 | 2904 | .950 |

[a]IPAQ: International Physical Activity Questionnaire.

[b]MET: metabolic equivalent of task.

For IPAQ sitting MET (Q7), the required sample sizes are relatively large since the correlation values are small. Thus, we can say that if we collect more data from a larger sample, the correlation between the paper survey and app survey of IPAQ Q7 can be significant. Even though the insignificance of correlation is due to the small sample size, the correlation values are also small (.061, .090, and .155), and thus we can say that the IPAQ Q7 shows different patterns in the paper-based survey and app survey. Figure 4 shows a graph of patient survey results with the largest difference in the values among those who surveyed more than once in 15 days, considering a week before and after the paper survey. There were increased or decreased trends in the app values, and there could be a gap between the values in the paper survey (single point) and the values in the app survey (the average of multiple points).

**Figure 4.** Examples of app data for (A) Distress Thermometer, (B) International Physical Activity Questionnaire Q2-6, and (C) International Physical Activity Questionnaire Q7.




Concordance correlation analysis, represented by the value C_b, in Table 4 also shows similar patterns with the previous correlation analysis. For DT and IPAQ (Q2-6), the C_b is larger than .90, which means that the values from the app-based survey are at least moderately matched with the paper-based survey if the values of app survey are recorded within 3 days before and after the paper survey. By contrast, for IPAQ (Q7), the C_b is smaller than .90, which means a poor association with the paper-based survey regardless of the time range of app inputs. Thus, we found that the app-based survey can substitute for the paper-based survey for DT and IPAQ (Q2-6), while the association between app-based inputs and paper survey of sitting time (IPAQ Q7) shows a poor association.

In Table 5, we show the coefficients on the value in the app survey and the odds ratio using ordered logistic regression. Like the results in the previous correlation analysis, for DT and IPAQ

total physical activity MET (Q2-6), the coefficients on the value in the app survey are all significant ($P$<.001), while the coefficients on app values for IPAQ sitting MET (Q7) are not significant for 3-day and 7-day intervals.

**Table 4.** Concordance correlation coefficients C_b between the value in the paper survey and the average value in the weekly app survey.

| Survey | Days before and after the paper survey and number of patients reporting through the app | | |
|---|---|---|---|
| | Same day (n=34) | 3 days (n=73) | 7 days (n=102) |
| Distress thermometer | .962 | .901 | .870 |
| IPAQ[a] total physical activity MET[b] (Q2-6) | .988 | .936 | .776 |
| IPAQ sitting MET (Q7) | .688 | .823 | .799 |

[a]IPAQ: International Physical Activity Questionnaire.

[b]MET: metabolic equivalent of task.

**Table 5.** Association between the value in the paper survey and the average value in the weekly app survey.

| | Average value in the app survey | | | | AIC[a] | Observations |
|---|---|---|---|---|---|---|
| | Beta | OR[b] | Z value | *P* value | | |
| **Distress thermometer** | | | | | | |
| Same day | 0.812 | 2.252 | 4.121 | <.001 | 154.95 | 34 |
| 3 days | 0.472 | 1.603 | 4.230 | <.001 | 366.73 | 73 |
| 7 days | 0.513 | 1.670 | 4.775 | <.001 | 498.66 | 102 |
| **IPAQ[c] total physical activity MET[d]** | | | | | | |
| Same day | 0.457 | 1.579 | 3.384 | <.001 | 239.47 | 34 |
| 3 days | 0.324 | 1.383 | 4.753 | <.001 | 533.03 | 73 |
| 7 days | 0.304 | 1.355 | 5.086 | <.001 | 783.43 | 102 |
| **IPAQ sitting MET** | | | | | | |
| Same day | 0.003 | 1.003 | 2.631 | .009 | 182.56 | 34 |
| 3 days | 0.001 | 1.001 | 1.296 | .20 | 412.78 | 73 |
| 7 days | 0.001 | 1.001 | 1.363 | .17 | 577.75 | 102 |

[a]AIC: Akaike information criterion.

[b]OR: odds ratio.

[c]IPAQ: International Physical Activity Questionnaire.

[d]MET: metabolic equivalent of task.

Table 6 shows the relationship between the answers of PHQ-9 (Q1-9) and PHQ (Q10) paper-based questionnaires and the answers of the corresponding app-based questionnaires. We analyzed the association between the paper-based questionnaires and the cases wherein the app-based questionnaires, answered on the same day when the paper-based questionnaire was answered or within 3, 7, or 14 days from the day. The PHQ-9 (Q1-9) questionnaires are standard format as shown in Figure 3C. In addition to PHQ-9 (Q1-9), PHQ (Q10) asks “if you checked off any problems among PHQ-9 (Q1-9), how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?”

According to the time interval between the paper-based questionnaire and the app-based questionnaire, the Spearman correlation values of the PHQ-9 sum value were .469 for 14 days, .574 for 7 days, .593 for 3 days, and .512 for the same day. Similarly, the correlation values for the answer to question 10 were .283 within 14 days, .409 within 7 days, .415 within 3 days, and .736 on the same day. Roughly speaking, the shorter the interval between the paper-based questionnaire and the app-based questionnaire, the higher the correlation value. All *P* values were statistically significant (Table 6).

**Table 6.** Spearman correlations between the values on the paper survey and the average values on the biweekly app survey.

| Survey | Days before and after the paper survey, number of patients reporting through the app, and scores | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Same day (n=18) | | | 3 days (n=39) | | | 7 days (n=63) | | | 14 days (n=96) | | |
| | Correlation | S-statistic | *P* value | Correlation | S-statistic | *P* value | Correlation | S-statistic | *P* value | Correlation | S-statistic | *P* value |
| PHQ-9[a] sum | .512 | 472.93 | .03 | .593 | 3720.6 | <.001 | .574 | 17,767 | <.001 | .469 | 78,362 | <.001 |
| PHQ-10[b] | .736 | 256.09 | .001 | .415 | 5345.6 | .009 | .409 | 24,640 | .001 | .283 | 105,715 | .005 |

[a]PHQ-9: Patient Health Questionnaire–9.

[b]PHQ-10: Patient Health Questionnaire Q10.

In Table 7, the required sample sizes are close to our sample size or slightly larger than our sample size. Even though some of our sample sizes were slightly smaller than the required sample sizes, the correlation coefficients were statistically significant. In addition, the actual powers are also sufficient as larger than .95.

In Table 8, we found that the app-based survey highly associated with the paper-based survey for the PHQ-9 sum value and PHQ (Q10), as the values of CCC are either close to .90 or larger than .90.

In Table 9, we show the coefficients on the value in the biweekly app survey and the odds ratio using ordered logistic regression. Like the results in the previous correlation analysis, for PHQ-9 sum and PHQ (Q10), the coefficients on the value in the app survey are all significant ($P<.05$). For PHQ-9 sum, the coefficients are smaller than 1, which means that the values in the app survey tend to be larger than the values in the paper survey. By contrast, for PHQ (Q10), the coefficients are larger than 1, which means that the values in the app survey tend to be smaller than the values in the paper survey.

**Table 7.** Sample size and actual statistical power in the correlation analysis of biweekly app survey.

| Survey | Days before and after the paper survey and scores | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Same day | | 3 days | | 7 days | | 14 days | |
| | Sample size | Actual power | Sample size | Actual power | Sample size | Actual power | Sample size | Actual power |
| PHQ-9[a] sum | 36 | .951 | 26 | .955 | 28 | .954 | 44 | .951 |
| PHQ-10[b] | 15 | .960 | 58 | .952 | 60 | .952 | 130 | .950 |

[a]PHQ-9: Patient Health Questionnaire–9.

[b]PHQ-10: Patient Health Questionnaire Q10.

**Table 8.** Concordance correlation coefficients C_b between the value in the paper survey and the average value in the biweekly app survey.

| Survey | Days before and after the paper survey, number of patients reporting through the app, and scores | | | |
|---|---|---|---|---|
| | Same day (n=18) | 3 days (n=39) | 7 days (n=63) | 14 days (n=96) |
| PHQ-9[a] sum | .941 | .967 | .901 | .894 |
| PHQ-10[b] | .940 | .938 | .879 | .936 |

[a]PHQ-9: Patient Health Questionnaire–9.

[b]PHQ-10: Patient Health Questionnaire Q10.

**Table 9.** Association between the value in the paper survey and the average value in the biweekly app survey.

| Survey | Average value in the app survey | | | | AIC[a] | Observations |
|---|---|---|---|---|---|---|
| | Beta | OR[b] | Z value | *P* value | | |
| **PHQ-9[c] sum** | | | | | | |
| Same day | 0.437 | 1.548 | 2.504 | .01 | 25.12 | 18 |
| 3 days | 0.296 | 1.344 | 3.655 | <.001 | 192.35 | 39 |
| 7 days | 0.282 | 1.326 | 4.402 | <.001 | 360.71 | 63 |
| 14 days | 0.247 | 1.280 | 5.253 | <.001 | 544.75 | 96 |
| **PHQ-10** [d] | | | | | | |
| Same day | 2.069 | 7.917 | 2.556 | .01 | 20.02 | 18 |
| 3 days | 2.168 | 8.741 | 2.885 | .004 | 56.65 | 39 |
| 7 days | 2.155 | 8.628 | 3.433 | .001 | 103.13 | 63 |
| 14 days | 1.343 | 3.831 | 3.679 | <.001 | 169.69 | 96 |

[a]AIC: Akaike information criterion.

[b]OR: odds ratio.

[c]PHQ-9: Patient Health Questionnaire–9.

[d]PHQ-10: Patient Health Questionnaire Q10.

# Discussion

## Principal Findings

As we are now able to diagnose and treat cancer early, the number of cancer survivors has increased worldwide. As the number of cancer survivors is increasing, we should pay more attention to them. Distress screening is a particularly important screening test for cancer survivors. According to the statistics of breast cancer patients, the prevalence of depression and anxiety among breast cancer survivors is 22% and 10%, respectively [16]. However, it is not easy for a clinician to diagnose a patient's stress early. Most cancer specialists do not have enough time to see their patients as they lack resources such as manpower, finances, and time to screen for pain, anxiety, or depression. In recent psychiatric research, a digital footprint created passively through mobile technology has been used as a tool for remote monitoring of patients. The feasibility of using data collected from mobile devices for developing a new measure of mental health has been established in several studies [17].

The most important aspect in collecting clinical data is the speed and reliability of data collection. Mobile technologies, which are widely used around the world, are also widely used in medical field and health care because they add a significant positive aspect of cost-effectiveness in managing data and improving the quality of clinical results [18]. Mobile health is a "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" as reported by the World Health Organization. If an electronic questionnaire is used instead of a paper-based questionnaire, the robustness of the electronic questionnaire must be established according to international guidelines first. This means that an equivalent measurement robustness should be demonstrated between both questionnaires [19].

The purpose of this study was to develop an electronic version of an original paper-based questionnaire and validate it. The electronic questionnaire retains the questions of the original paper-based questionnaire; however, the layout has been adapted for use on smartphones. The app was developed keeping in mind the ease of use not only for patients who have experience with smartphones but also for those who are new to smartphones. In the process of making a paper-based questionnaire into an electronic questionnaire, it is unclear which correction is being referred to. Because of this correction process, equivalence analysis between the electronic questionnaire and paper-based questionnaire needs to be conducted. To create the app used in this study, we needed to adjust the ePRO appropriately. We adjusted the size of the text and added a ScrollView feature that enables scrolling down to see all the sentences. This calibration process is why equivalence analysis is necessary. No significant difference was found between the two approaches for all the items investigated. This result corroborates previous studies stating that there were few or no differences between the electronic and paper-based questionnaires [20,21]. In a previous study, Bierbrier et al [22] did not focus on measuring instruments related to the musculoskeletal system but experimented with various mHealth apps from the Google Play Store and App Store to assess the accuracy of the electronic physiotherapy questionnaire using a smartphone.

It is important that the data obtained from the app is accurate. Therefore, the possibility of errors due to human involvement should be eliminated. There is some evidence to support the promising role of a mobile app to remotely collect PRO, as the mobile app-based PRO questionnaires could be free from generating a bias driven by Hawthorne effect. Bush et al [23]

evaluated active military personnel in the United States and administered a 7-question questionnaire via an app and on paper, and the responses were similar. Garcia-Palacios et al [24] have also conducted a study on the use of questionnaires in patients with fibromyalgia. In this study, the mean pain and fatigue scores reported in paper and smartphone questionnaires were not statistically significantly different. ePRO for patients has been found to be in good agreement with the results of the paper version [18]. In a study on the validity and reliability of International Prostate Symptom Score, Kim et al [25] reported a strong correlation between results of smartphone and paper-based questionnaires.

It is important to be able to receive useful information data through ePRO. It is possible to collect data several times a day whenever the patient wants. In this way, we can collect continuous data over time and use this database to create a more strategic and systematic treatment plan for doctors to treat cancer survivors. In addition, it not only benefits patients but also significantly changes and advances the overall health and health care system [18].

When using questionnaires to assess health outcomes by electronic methods can provide more accurate and efficient information than paper-based methods, and since the information is communicated through mobile phones, the answers can be safely achieved. This suggests that patients may be better served electronically than with paper [18]. It has also been reported that responding to a questionnaire using a mobile phone is faster and safer than it is using a paper-based questionnaire [25,26].

These results suggest that mobile devices have a significant potential as tools for distress screening in patients with unmet needs. In agreement with previous reports, during the 6-month study of 106 patients who participated in this study, breast cancer patients collaborated very well with data collection using smartphone apps and maintained a high level of compliance. Data for the DT questionnaire, which is answered every week, and the PHQ-9, which is answered every 2 weeks, were collected through the same app, and overall collection rates were 42.42% (3597/8480) and 41.86% (1775/4240), respectively. This suggests that ePRO has a greater effect on improving patient lifestyle than paper or web-based methods [27].

In our study, the Spearman correlation between the paper-based questionnaire and app-based questionnaire is distributed between .30 and .80 depending on the answer interval, except the IPAQ (Q7; sitting MET). The shorter the time interval was, the larger the value.

Even in the same patient, since emotional factors such as distress change frequently, DT (Figure 4A) also changed depending on the answer time. Overall, although the paper-based questionnaire is admittedly a standard test, a smartphone app-based questionnaire can be a better way to identify patient conditions that change from time to time and incorporate the findings into a treatment plan.

## Limitations

Our research has some limitations. First, we did not evaluate the experience and familiarity with smartphone use of patients in the study. A previous study has reported that patients found completing the ePRO questionnaire more comfortable and better than completing the paper-based questionnaire [19]. We also did not consider the time it takes to complete the survey. A previous study has reported that answering a paper-based questionnaire takes less time than answering an app-based questionnaire [19].

In addition, all subjects in this study were Android users. Thus, one may have a concern on the sample selection bias driven by the focus on an Android app. However, iPhone penetration in the elderly in Korea is not very great, thus the selection bias is not serious in this study. Finally, the subjects in this study are younger than the total breast cancer patient population as we recruited a sample that could use a smartphone. Thus, the effectiveness of a mobile app to collect patient-reported information for older patients should be validated in a future study.

Despite these limitations, our study is the first empirical effort to demonstrate the effectiveness of a smartphone app-based version of a questionnaire associated with distress and physical activity using a proper statistical analysis. Using electronic methods to measure the PROs of distress and physical activity can help doctors and patients to more effectively collect the information. Doctors can evaluate distress and physical activity, improve their clinical observation, and make better decisions related to treatment at the time of patient consultation. In addition, there is a growing interest in patient-centered care, which shows that patients participate in their health care and improve their health. Therefore, having the opportunity to manage distress and physical activity at home, and evaluate the results with doctors has the advantage of strengthening the relationship between patients and doctors.

## Conclusions

The app-based questionnaire related to stress and physical activity is a useful assessment tool for health care professionals. Using the app, clinicians can easily collect answers to questionnaires from their patients and effectively manage, store, and organize them. The app shows a high level of validity and compliance for DT, IPAQ (Q2-6), and PHQ-9. In conclusion, the app version of the questionnaire was not inferior to the paper version, and there was sufficient potential for equal use in clinical practice.

## Acknowledgments

This study was supported by a grant from the National R&D Program for Cancer Control, Ministry of Health and Welfare, Republic of Korea (HA17C0047), the Korean Breast Cancer Foundation, Seoul, Republic of Korea (KBCF-2015E003), and the Asan Institute for Life Sciences and Corporate Relations of Asan Medical Center, Seoul, Republic of Korea (2017-0328).

## Authors' Contributions

IYC, JWL, and SC designed the study and provided input throughout the study. YRP, HC, HJP, and ML collected the data. JK, HJK, BSK, SC, BHS, and SHA provided clinical expertise throughout the study. MJ, DC, SBL, and IYC analyzed portions of the data. SBL and MJ wrote the manuscript along with contributions from all authors. All authors have read and approved the final manuscript.

## Conflicts of Interest

HC is CEO of Swallaby Co Ltd, Seoul, Korea.

Multimedia Appendix 1
Korean language version of the (A) Distress Thermometer, (B) International Physical Activity Questionnaire, and (C) Patient Health Questionnaire–9.
[PNG File , 1191 KB - mhealth_v8i5e17320_app1.png ]

## References


1. Kim J, Lim S, Min YH, Shin Y, Lee B, Sohn G, et al. Depression screening using daily mental-health ratings from a smartphone application for breast cancer patients. J Med Internet Res 2016 Aug 04;18(8):e216 [FREE Full text] [doi: 10.2196/jmir.5598] [Medline: 27492880]
2. Mitchell AJ, Ferguson DW, Gill J, Paul J, Symonds P. Depression and anxiety in long-term cancer survivors compared with spouses and healthy controls: a systematic review and meta-analysis. Lancet Oncol 2013 Jul;14(8):721-732. [doi: 10.1016/S1470-2045(13)70244-4] [Medline: 23759376]
3. FDA Center for Drug Evaluation and Research, FDA Center for Biologics Evaluation and Research, FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79 [FREE Full text] [doi: 10.1186/1477-7525-4-79] [Medline: 17034633]
4. Jette DU, Halbert J, Iverson C, Miceli E, Shah P. Use of standardized outcome measures in physical therapist practice: perceptions and applications. Phys Ther 2009 Feb;89(2):125-135. [doi: 10.2522/ptj.20080234] [Medline: 19074618]
5. Hatfield DR, Ogles BM. Why some clinicians use outcome measures and others do not. Adm Policy Ment Health 2007 May;34(3):283-291. [doi: 10.1007/s10488-006-0110-y] [Medline: 17211715]
6. Schick-Makaroff K, Molzahn A. Strategies to use tablet computers for collection of electronic patient-reported outcomes. Health Qual Life Outcomes 2015;13:2 [FREE Full text] [doi: 10.1186/s12955-014-0205-1] [Medline: 25608560]
7. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
8. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
9. 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]
10. Duregger K, Hayn D, Nitzlnader M, Kropf M, Falgenhauer M, Ladenstein R, et al. Electronic patient reported outcomes in paediatric oncology: applying mobile and near field communication technology. Stud Health Technol Inform 2016;223:281-288. [Medline: 27139415]
11. Burdette SD, Herchline TE, Oehler R. Surfing the web: practicing medicine in a technological age: using smartphones in clinical practice. Clin Infect Dis 2008 Jul 01;47(1):117-122. [doi: 10.1086/588788] [Medline: 18491969]
12. Juniper EF, Langlands JM, Juniper BA. Patients may respond differently to paper and electronic versions of the same questionnaires. Respir Med 2009 Jun;103(6):932-934 [FREE Full text] [doi: 10.1016/j.rmed.2008.10.019] [Medline: 19028085]
13. Zbrozek A, Hebert J, Gogates G, Thorell R, Dell C, Molsen E, et al. Validation of electronic systems to collect patient-reported outcome (PRO) data—recommendations for clinical trial teams: report of the ISPOR ePRO systems validation good research practices task force. Value Health 2013 Jun;16(4):480-489 [FREE Full text] [doi: 10.1016/j.jval.2013.04.002] [Medline: 23796281]
14. Swallaby Inc. URL: http://www.swallaby.com/ [accessed 2020-04-20]
15. Lin LI. Assay validation using the concordance correlation coefficient. Biometrics 1992 Jun;48(2):599-604. [doi: 10.2307/2532314]
16. Zainal NZ, Nik-Jaafar NR, Baharudin A, Sabki ZA, Ng CG. Prevalence of depression in breast cancer survivors: a systematic review of observational studies. Asian Pac J Cancer Prev 2013;14(4):2649-2656 [FREE Full text] [Medline: 23725190]
17. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]

18. Epis OM, Casu C, Belloli L, Schito E, Filippini D, Muscarà M, et al. Pixel or paper? Validation of a mobile technology for collecting patient-reported outcomes in rheumatoid arthritis. JMIR Res Protoc 2016 Nov 16;5(4):e219 [FREE Full text] [doi: 10.2196/resprot.5631] [Medline: 27852561]
19. Duracinsky M, Lalanne C, Goujard C, Herrmann S, Cheung-Lung C, Brosseau J, et al. Electronic versus paper-based assessment of health-related quality of life specific to HIV disease: reliability study of the PROQOL-HIV questionnaire. J Med Internet Res 2014;16(4):e115 [FREE Full text] [doi: 10.2196/jmir.3330] [Medline: 24769643]
20. Thorén ES, Andersson G, Lunner T. The use of research questionnaires with hearing impaired adults: online vs. paper-and-pencil administration. BMC Ear Nose Throat Disord 2012;12:12 [FREE Full text] [doi: 10.1186/1472-6815-12-12] [Medline: 23107440]
21. Holländare F, Andersson G, Engström I. A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. J Med Internet Res 2010;12(5):e49 [FREE Full text] [doi: 10.2196/jmir.1392] [Medline: 21169165]
22. Bierbrier R, Lo V, Wu RC. Evaluation of the accuracy of smartphone medical calculation apps. J Med Internet Res 2014;16(2):e32 [FREE Full text] [doi: 10.2196/jmir.3062] [Medline: 24491911]
23. Bush NE, Skopp N, Smolenski D, Crumpton R, Fairall J. Behavioral screening measures delivered with a smartphone app: psychometric properties and user preference. J Nerv Ment Dis 2013 Nov;201(11):991-995. [doi: 10.1097/NMD.0000000000000039] [Medline: 24177488]
24. Garcia-Palacios A, Herrero R, Belmonte MA, Castilla D, Guixeres J, Molinari G, et al. Ecological momentary assessment for chronic pain in fibromyalgia using a smartphone: a randomized crossover study. Eur J Pain 2014 Jul;18(6):862-872. [Medline: 24921074]
25. Kim JH, Kwon S, Shim SR, Sun HY, Ko YM, Chun D, et al. Validation and reliability of a smartphone application for the International Prostate Symptom Score questionnaire: a randomized repeated measures crossover study. J Med Internet Res 2014 Feb 10;16(2):e38 [FREE Full text] [doi: 10.2196/jmir.3042] [Medline: 24513507]
26. Salaffi F, Gasparini S, Ciapetti A, Gutierrez M, Grassi W. Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. Rheumatology (Oxford) 2013 Nov;52(11):2062-2070 [FREE Full text] [doi: 10.1093/rheumatology/ket276] [Medline: 23955646]
27. Chung IY, Jung M, Lee SB, Lee JW, Park YR, Cho D, et al. An assessment of physical activity data collected via a smartphone app and a smart band in breast cancer survivors: observational study. J Med Internet Res 2019 Sep 06;21(9):13463 [FREE Full text] [doi: 10.2196/13463] [Medline: 31493319]


## Abbreviations

**CCC:** concordance correlation coefficient
**DT:** Distress Thermometer
**ePRO:** electronic patient-reported outcome
**IPAQ:** International Physical Activity Questionnaire
**MET:** metabolic equivalent of task
**mHealth:** mobile health
**PHQ-9:** Patient Health Questionnaire–9
**PRO:** patient-reported outcome


*Edited by G Eysenbach; submitted 08.12.19; peer-reviewed by Y Chu, H Chen, R Haghighi Osgouei; comments to author 26.02.20; revised version received 05.03.20; accepted 15.04.20; published 04.05.20.*

*Please cite as:*
*Jung M, Lee S, Kim J, Kim H, Ko B, Son BH, Ahn SH, Park YR, Cho D, Chung H, Park HJ, Lee M, Lee JW, Chung S, Chung IY*
*A Mobile Technology for Collecting Patient-Reported Physical Activity and Distress Outcomes: Cross-Sectional Cohort Study*
*JMIR Mhealth Uhealth 2020;8(5):e17320*
*URL: https://mhealth.jmir.org/2020/5/e17320*
*doi:10.2196/17320*
*PMID:32364508*

Original Paper

# A Mobile Health App for the Collection of Functional Outcomes After Inpatient Stroke Rehabilitation: Pilot Randomized Controlled Trial

Li Li[1*], MS; Jia Huang[1,2*], PhD; Jingsong Wu[1], PhD; Cai Jiang[1], PhD; Shanjia Chen[1], PhD; Guanli Xie[1], PhD; Jinxin Ren[1], MS; Jing Tao[1,2,3], PhD; Chetwyn C H Chan[4], PhD; Lidian Chen[1,2,3], PhD; Alex W K Wong[5,6,7], PhD, DPhil

[1]College of Rehabilitation Medicine, Fujian University of Traditional Chinese Medicine, Fuzhou, China

[2]Key Laboratory of Orthopedics & Traumatology of Traditional Chinese Medicine and Rehabilitation, Fujian University of Traditional Chinese Medicine, Ministry of Education, Fuzhou, China

[3]Traditional Chinese Medicine Rehabilitation Research Center of State Administration of Traditional Chinese Medicine, Fuzhou, China

[4]Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, HongKong, Hong Kong

[5]Program in Occupational Therapy, Washington University School of Medicine, St. Louis, MO, United States

[6]Department of Neurology, Washington University School of Medicine, St. Louis, MO, United States

[7]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, United States

[*]these authors contributed equally

**Corresponding Author:**
Alex W K Wong, PhD, DPhil
Program in Occupational Therapy
Washington University School of Medicine
600 South Taylor Avenue, MSC 8505-94-01
St. Louis, MO
United States
Phone: 1 314 286 0278
Email: wongal@wustl.edu

## Abstract

**Background:** Monitoring the functional status of poststroke patients after they transition home is significant for rehabilitation. Mobile health (mHealth) technologies may provide an opportunity to reach and follow patients post discharge. However, the feasibility and validity of functional assessments administered by mHealth technologies are unknown.

**Objective:** This study aimed to evaluate the feasibility, validity, and reliability of functional assessments administered through the videoconference function of a mobile phone–based app compared with administration through the telephone function in poststroke patients after rehabilitation hospitalization.

**Methods:** A randomized controlled trial was conducted in a rehabilitation hospital in Southeast China. Participants were randomly assigned to either a videoconference follow-up (n=60) or a telephone follow-up (n=60) group. We measured the functional status of participants in each group at 2-week and 3-month follow-up periods. Half the participants in each group were followed by face-to-face home visit assessments as the gold standard. Validity was assessed by comparing any score differences between videoconference follow-up and home visit assessments, as well as telephone follow-up and home visit assessments. Reliability was assessed by computing agreements between videoconference follow-up and home visit assessments, as well as telephone follow-up and home visit assessments. Feasibility was evaluated by the levels of completion, satisfaction, comfort, and confidence in the 2 groups.

**Results:** Scores obtained from the videoconference follow-up were similar to those of the home visit assessment. However, most scores collected from telephone administration were higher than those of the home visit assessment. The agreement between videoconference follow-up and home visit assessments was higher than that between telephone follow-up and home visit assessments at all follow-up periods. In the telephone follow-up group, completion rates were 95% and 82% at 2-week and 3-month follow-up points, respectively. In the videoconference follow-up group, completion rates were 95% and 80% at 2-week and 3-month follow-up points, respectively. There were no differences in the completion rates between the 2 groups at all follow-up periods ($X^2_1$=1.6, $P$=.21 for 2-week follow-up; $X^2_1$=1.9, $P$=.17 for 3-month follow-up). Patients in the videoconference follow-up

group perceived higher confidence than those in the telephone follow-up group at both 2-week and 3-month follow-up periods ($X^2_3$=6.7, *P*=.04 for 2-week follow-up; $X^2_3$=8.0, *P*=.04 for 3-month follow-up). The videoconference follow-up group demonstrated higher satisfaction than the telephone follow-up group at 3-month follow-up ($X^2_3$=13.9; *P*=.03).

**Conclusions:** The videoconference follow-up assessment of functional status demonstrates higher validity and reliability, as well as higher confidence and satisfaction perceived by patients, than the telephone assessment. The videoconference assessment provides an efficient means of assessing functional outcomes of patients after hospital discharge. This method provides a novel solution for clinical trials requiring longitudinal assessments.

**Trial Registration:** chictr.org.cn: ChiCTR1900027626; http://www.chictr.org.cn/edit.aspx?pid=44831&htm=4.

***(JMIR Mhealth Uhealth 2020;8(5):e17219)*** doi:10.2196/17219

**KEYWORDS**

telemedicine; cell phone; stroke; rehabilitation; activities of daily living; outcome and process assessment; health care

## Introduction

Stroke is one of the leading causes of death and disability worldwide [1]. As in other developing countries, the incidence and prevalence of stroke in China are gradually increasing. Each year, China has 2.5 million new stroke cases and more than 11 million stroke survivors; stroke has become the leading cause of death in China [2]. Stroke can have a long-term impact on an individual's physical, mental, and social function, as well as on survivors' caregivers and families [3-6]. The majority of patients receive inpatient stroke rehabilitation to regain function for only the first few weeks after stroke, but functional recovery often occurs 3 months or even longer following a stroke [7]. Additionally, about 30% of poststroke individuals receive outpatient rehabilitation [8]. Even if available, access to rehabilitation for patients in China is limited because of transportation, geographical barriers, and monetary factors [9]. As most patients have recovery potential but do not receive recommended rehabilitation, it is important to develop new strategies to continuously monitor functional recovery and other health outcomes of patients following discharge to better understand the long-term consequences for patients poststroke [10].

Poststroke home-based therapies seem to be a viable option for the delivery of stroke care. Follow-up assessments and interventions not only provide a means of monitoring the functional status of patients after transitioning to home and community [11], but also provide instructions to prevent readmission, which is especially important for those who receive a longer stay in inpatient rehabilitation [12]. Moreover, follow-up assessments enable clinicians to adjust the treatment plan for home-based therapies [13,14]. One common method for follow-up assessment is a face-to-face, at-home assessment in which the home health therapist visits the patient at home to perform an evaluation. However, this method demands intensive resources, including the time of trained personnel and financial expenditures [15]. Recent studies have tested alternative methods of follow-up data collection for patients following a stroke [16-18].

Telephone administration is a common alternative. This method allows participants to be recruited from diverse geographical areas, is typically less expensive than the face-to-face home assessment, and has a quick turnaround time [19]. Prior studies have found that telephone administration of outcome measures demonstrates equivalent reliability to face-to-face assessment, supporting telephone interview as a feasible solution [20,21]. However, there are some shortcomings to telephone administration. First, many functional measures require a trained therapist to observe and provide ratings of how the patient performs specific daily tasks. Inability to perform observations via telephone administration may be a hindrance to accurately evaluating task performance. Furthermore, telephone administration often assesses survey-based questions, which require patients to have higher education, health literacy, and communication abilities to understand the verbal instructions [22]. An earlier study found a large amount of missing data from assessments administered through the telephone interview method for stroke patients and caregivers, limiting the use of this method in clinical trials requiring longitudinal assessments [23].

With advances in computing power and mobile connectivity, many mobile health (mHealth) technologies, such as mobile devices, sensors, apps, and social media, are becoming available to obtain data pertinent to wellness and disease diagnosis, prevention, and management [24]. WeChat (Chinese version: Weixin), developed in 2011 by Tencent, has become the most common social software app in China [25]. Similar to other social media apps, such as Facebook, Twitter, and WhatsApp, WeChat is a free platform that provides seamless opportunities for communication and other mobile apps. People can communicate with one another through the free voice call or video call feature, as well as instantly share information [26]. According to the Statista Research Department [27], WeChat had over 1.15 billion monthly active users from a wide range of age groups. Harnessing the use of mHealth technologies to improve health and wellness is not uncommon in modern health care [28]. A recent mHealth intervention study utilizing the WeChat app for weight loss behaviors in a group of male workers found promising results: participants who spent more time using the health education program embedded in the WeChat app for engaging in healthy behaviors demonstrated more weight loss [29]. Another mHealth study used the WeChat app to educate parents of pediatric patients undergoing surgery and found that this mobile app–assisted intervention was effective in enhancing parents' knowledge of perioperative procedures [30]. Another study used the WeChat app in a group of discharged patients with head and neck tumors for 6 months

and demonstrated the app to be a cost-effective method of follow-up assessment [31].

Although the utility of the WeChat app has been demonstrated in other populations, little is known about whether the WeChat app could be feasible to assess the postdischarge functional status of patients following a stroke, because most poststroke patients experience cognitive and communication difficulties that may make it difficult to operate the app and understand its instructions. To address this question, we conducted a pilot randomized controlled trial in a group of discharged stroke patients by randomly assigning them into 2 different modes of administration during 2-week and 3-month follow-up periods: WeChat video conference or WeChat telephone administration. This study had 2 specific aims. The first aim was to compare the validity and reliability of functional assessment between these 2 modes of administration in stroke patients. We hypothesized that videoconference administration would demonstrate higher validity and reliability than telephone administration, because examiners using videoconference administration can observe how the respondent performs specific tasks to provide appropriate ratings, whereas examiners in the telephone administration group demand more subjective appraisals of task performance based on the respondent's verbal descriptions. Our second aim was to examine the feasibility of the functional assessment administered via the videoconference function compared with the telephone call function in stroke patients after rehabilitation hospitalization. We hypothesized that both modes of administration would demonstrate high levels of completion, comfort, satisfaction, and confidence.

## Methods

### Study Design

This study was a parallel, 2-group, and pragmatic randomized controlled trial of an mHealth app of functional outcome data collection after inpatient stroke rehabilitation. The trial was registered with the Chinese Clinical Trial Registry: ChiCTR1900027626. A total of 120 eligible stroke patients from the affiliated rehabilitation hospital of Fujian University of Traditional Chinese Medicine were recruited for this study. Participants were approached by a research assistant, who provided the study information. After participants provided informed consent and were screened for eligibility, participants were randomized into 1 of the 2 WeChat app administration groups: videoconference follow-up or telephone follow-up, with a ratio of 1:1. Eligible participants were randomized using a random number table generated by a study coordinator who was not involved in the recruitment and assessment of participants for the study.

### Ethics Approval

This trial was implemented in compliance with the declaration of Helsinki and approved by the Ethics Board of the affiliated rehabilitation hospital of Fujian University of Traditional Chinese Medicine (number: 2016KY-032-01). All participants provided informed, written consent before participation. Participants received an honorarium to acknowledge their research contribution.

### Participants

Patients were eligible to participate in the study if they met the following criteria: (1) aged at least 18 years, (2) diagnosis of first stroke, (3) normal speech function according to the Mandarin Language Screening Test with cutoff scores of >13 for those with primary school education or >14 for those with junior high school or higher education, (4) normal cognitive function according to the Montreal Cognitive Assessment with a cutoff score of >26, and (5) home discharge. Participants were excluded if they (1) did not own a mobile phone, (2) were unwilling to install and use the WeChat software on their mobile phone, (3) had emotional dysfunction according to the Beck Depression Inventory with a cutoff score of >13, or (4) had other medical illnesses limiting study participation. As we included only participants who served as their own informant rather than including participants on a nonselected basis, it is likely that individuals who were too cognitively impaired or were unable to understand the study materials were excluded.

### Recruitment and Screening

Participants were recruited from the inpatient rehabilitation hospital. Recruitment was initiated while the stroke patient was still in the hospital. The research assistant screened the medical records of all patients undergoing inpatient rehabilitation for a stroke. Once potential participants were identified, the research assistant approached the individual and provided information about the study. Participants provided informed consent once they agreed to participate. The research assistant reviewed the inclusion and exclusion criteria and completed the screening tests to confirm eligibility. Participants were then enrolled, and randomization occurred only after recruitment by a study coordinator.

### Data Collection Procedures

All research assistants received training in assessing the eligibility of potential participants, obtaining informed consent from participants, the study protocol, and obtaining outcome measures for both groups. They also received training in how to coach and assist participants using the WeChat app, including the videoconference and telephone call functions. After randomization, all participants received the baseline assessment at the week of discharge, followed by the completion of 2 mHealth app follow-up sessions (either videoconference or telephone), and half of the participants from each group received 2 home visits. The first follow-up session occurred 2 weeks after home discharge. Within 1 week of the first follow-up session, half of the study participants were selected to conduct the first home visit based on stratified sampling in each group. The stratified sampling criteria were grounded on participants' functional abilities. As home visits are costly, this study was limited by randomly selecting half of the study participants for the home visit assessment. The second follow-up session occurred 3 months after home discharge. Within 1 week of the second follow-up session, we completed the second home visit in this subgroup of participants. The time interval between videoconference or telephone follow-up and home visit of 1 week was considered long enough to ensure that the previous responses were forgotten and short enough to ensure that the patient's clinical condition would not substantially change. All

assessments were conducted by our trained research assistants with training in physical therapy, occupational therapy, or rehabilitation medicine. To reduce assessor bias in the trial, research assistants who completed the videoconference or telephone follow-up sessions also conducted home visits with patients in the same group. We treated the face-to-face, home visit assessment as the gold standard in this study.

Participants in the videoconference follow-up group received training on the usage of the videoconference function of the WeChat app before discharge. During the follow-up sessions, research assistants asked participants to complete individual functional tasks and rate their actual performance through the videoconference, with the exception of bladder and bowel management tasks, which were assessed by the participant's verbal descriptions. Participants also described difficulties pertaining to their individual task performance. During the home visit, research assistants completed the face-to-face observations by rating participants as they completed the same functional tasks. We used the same scoring criteria to evaluate the performance of our study participants in both videoconference and face-to-face, home visit assessments. Participants in the telephone follow-up group received training on the usage of the telephone function of the WeChat app before discharge. During the follow-up sessions, research assistants made telephone calls, asked participants how they performed in each functional item, and appraised their performance based on the participant's verbal descriptions. During the home visit, research assistants completed the same protocol as in the videoconference follow-up group. We used the same scoring criteria to evaluate the performance of our study participants in both telephone and home visit assessments.

### Outcome Measures

As recommended in a systematic review of optimal outcome measures for stroke therapy trials [32], the primary outcome selected for this study was improvement in activities and participation rather than the reduction of impairments. Thus, we chose the functional status (ie, the performance of activities of daily living, ADLs) of stroke participants as our primary outcome variable. We used the Modified Barthel Index (MBI) to assess the performance of ADLs in the 2 groups [33,34]. The MBI can be administered using clinician-rated or patient-reported methods. It includes 10 items measuring grooming, bathing, feeding, toileting, stair climbing, dressing, bowel control, bladder control, mobility, and chair/bed transfer. Items have different response options, with anchored scores provided for different options. The total score ranges from 0 to 100. A higher score means that the participant has greater independence. Adequate validity and reliability were found for the Chinese version of the MBI used in this study [35].

We also defined 2 outcome variables to examine the feasibility of using an mHealth app to measure functional status of stroke participants in this study. The first variable was the completion of the MBI among our study participants in both groups at both follow-up periods. The second variable was the acceptability among our study participants (ie, levels of satisfaction, comfort, and confidence) of using the videoconference or telephone functions to complete the functional assessment at follow-up periods. We developed 3 questions: (1) "Are you satisfied with this follow-up assessment?" (2) "Are you comfortable with this follow-up assessment?" and (3) "Are you confident using this follow-up assessment?" All items were rated on a 4-point scale ranging from "very satisfied/comfortable/confident" to "unsatisfied/uncomfortable/unconfident."

### Statistical Analysis

Data were analyzed using the Statistical Package for Social Sciences, version 20 (IBM, Chicago, IL, United States). Baseline characteristics between groups were compared using *t* tests or Mann-Whitney tests for continuous variables, and Pearson chi-square tests for categorical variables. We applied Wilcoxon signed-rank tests to compare MBI score differences between videoconference and telephone assessments, as well as between videoconference/telephone and home visit assessments for the validity evaluation. We set the *P* value to .05 for statistical significance. We used intraclass correlation coefficients (ICCs) to evaluate the agreement of all item scores between these assessments for the reliability evaluation. According to Landis and Koch [36], the ICC can theoretically vary between 0 and 1.0, where an ICC of 0 indicates no reliability, and an ICC of 1.0 indicates perfect reliability; ICCs above 0.80 indicate acceptable reliability. We used chi-square tests or Fisher exact tests to compare rates of completion, satisfaction, comfort, and confidence between the videoconference and telephone modes of administration.

## Results

### Baseline Characteristics

Figure 1 illustrates the flow of participant enrollment. Among 519 potential stroke inpatients, 353 did not meet the inclusion criteria. A total of 21 patients refused to participate in the study because family members were uncertain about the use of mHealth for collecting data. Some patients reported that they could easily access medical services and did not require additional follow-up services. In total, 25 patients were discharged from the hospital before research assistants approached them. A total of 120 participants were successfully recruited and randomized to 1 of the 2 groups. Table 1 describes the demographic characteristics of study participants. Study participants were middle-aged (mean age 59.7 years, SD 12.1), 59.1% (71/120) of participants were women, and 93.3% (112/120) of participants were married. In total, 45.8% (55/120) of participants completed 9 or fewer years of formal education, and 28.3% (34/120) of participants were currently employed. A total of 45.0% (54/120) of participants had a history of cerebral infarction (ie, ischemic stroke). We found no significant differences between the 2 groups on gender, marital status, education, occupation, type of stroke, or duration of disease. We also found no significant differences between the 2 groups in any functional task measured by the MBI at the time of discharge. Eight participants in the videoconference follow-up group dropped out at 2 weeks: 3 participants did not answer the video calls and 5 participants refused to complete the assessment. Three participants in the telephone follow-up group dropped out at 2 weeks; they did not answer the telephone calls. At the 3-month follow-up point, we lost more participants: 3

participants in the videoconference follow-up group and 9 in the telephone follow-up group. Of the initial 60 participants in each group, 82% (49/60) of participants in the videoconference follow-up group and 80% (48/60) of participants in the telephone follow-up group completed the entire study protocol.

**Figure 1.** A flow diagram of patient enrollment.

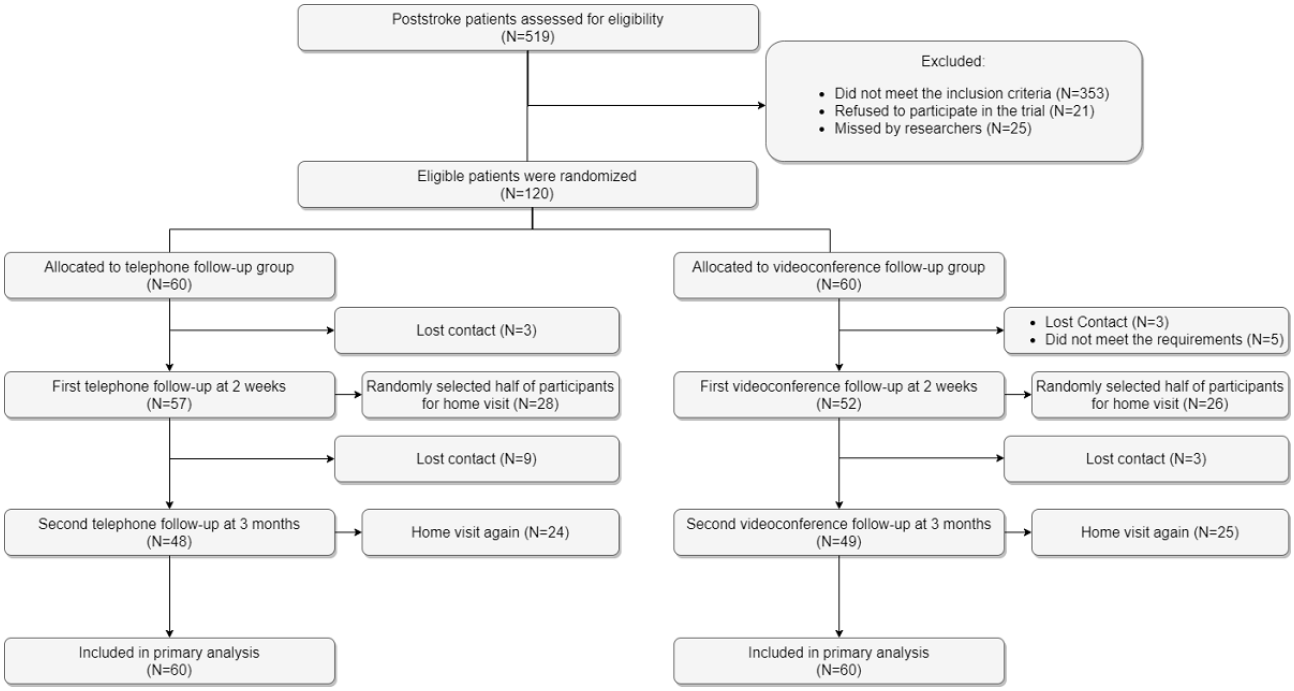


**Table 1.** Baseline characteristics of participants.

| Variables | All participants (N=120) | Telephone follow-up (n=60) | Videoconference follow-up (n=60) | *Z* test or $X^2$ (df) test | *P* value |
|---|---|---|---|---|---|
| Age (years), mean (SD) | 59.7 (12.1) | 60.8 (11.6) | 58.3 (12.4) | 1.56 | .25 |
| **Gender, n (%)** | | | | **0.14 (1)** | **.85** |
| Male | 49 (40.8) | 24 (41.7) | 25（36.7） | | |
| Female | 71 (59.1) | 36（58.3） | 35（63.3） | | |
| **Marital status, n (%)** | | | | **0.53 (1)** | **.72** |
| Married | 112 (93.3) | 57（95.0） | 55（91.7） | | |
| Other | 8 (6.7) | 3（5.0） | 5（8.3） | | |
| **Education (years), n (%)** | | | | **1.38 (3)** | **.71** |
| ≤6 | 25 (20.8) | 14（23.3） | 11（18.3） | | |
| 7-9 | 41 (34.2) | 22（36.7） | 19（31.7） | | |
| 9-12 | 30 (25.0) | 14（23.3） | 16（26.7） | | |
| ≥12 | 24 (20.0) | 10（16.7） | 14（23.3） | | |
| **Occupation, n (%)** | | | | **1.57 (3)** | **.67** |
| Employed | 34 (28.3) | 19（31.7） | 15（25.0） | | |
| Retired | 50 (41.7) | 26（43.3） | 24（40） | | |
| Unemployed | 10 (8.3) | 4（6.7） | 6（10） | | |
| Other | 26 (21.7) | 11（18.3） | 15（25） | | |
| **Type of stroke, n (%)** | | | | **2.15 (1)** | **.14** |
| Infarction | 54 (45.0) | 23（38.3） | 31（51.7） | | |
| Hemorrhage | 66 (55.0) | 37（61.7） | 29（48.3） | | |
| Duration of stroke (days), mean (SD) | 90.7 (13.8) | 87.6 (14.6) | 93.7 (12.9) | 0.53 | .24 |
| **Discharge functional status (Modified Barthel Index), mean (SD)** | | | | | |
| Feeding | 7.27 (2.09) | 7.35 (1.92) | 7.12 (2.08) | –0.59 | .55 |
| Grooming | 3.06 (0.83) | 3.02 (0.68) | 3.15 (0.90) | –1.64 | .10 |
| Dressing | 5.32 (2.10) | 5.08 (2.09) | 5.62 (2.27) | –1.35 | .18 |
| Bathing | 2.02 (1.46) | 2.23 (1.09) | 1.85 (1.15) | –2.02 | .06 |
| Toilet use | 5.21 (2.24) | 5.13 (1.82) | 5.32 (2.14) | –0.49 | .62 |
| Bowels | 8.87 (1.64) | 8.65 (1.63) | 9.08 (1.33) | –1.30 | .19 |
| Bladder | 9.06 (1.17) | 9.00 (1.24) | 9.18 (1.11) | –1.24 | .22 |
| Transfer | 8.57 (2.89) | 8.40 (2.68) | 8.70 (3.03) | –0.68 | .49 |
| Mobility | 7.73 (2.89) | 8.15 (2.92) | 7.42 (3.31) | –1.34 | .18 |
| Stairs | 4.66 (2.17) | 4.48 (1.87) | 4.83 (2.27) | –0.77 | .44 |

## Comparison of Videoconference and Telephone Follow-Up Assessments

Table 2 shows the MBI scores for videoconference follow-up and telephone follow-up at 2-week and 3-month periods. We found no significant differences between the 2 groups in the majority of functional tasks measured by the MBI at 2 weeks and 3 months, except that significant differences were found in the bladder management task at 2 weeks, and the grooming and bathing tasks at 3 months. ICC values for all but the grooming task at 2 weeks and the grooming, toilet use, and mobility tasks at 3 months exceeded 0.8, indicating acceptable reliability between videoconference and telephone assessments at both follow-up periods.

**Table 2.** Comparison of Modified Barthel Index scores evaluated by videoconference and telephone follow-up assessments at 2-week follow-up and 3-month follow-up.

| Variables | Telephone follow-up, mean (SD) | Videoconference follow-up, mean (SD) | *Z* test | *P* value | Intraclass correlation coefficient |
|---|---|---|---|---|---|
| **2-week follow-up** | | | | | |
| Feeding | 8.64 (1.59) | 8.44 (1.72) | –1.64 | .10 | 0.82 |
| Grooming | 3.65 (0.69) | 3.45 (0.76) | –1.79 | .08 | 0.74 |
| Dressing | 6.74 (2.00) | 6.44 (2.03) | –1.90 | .06 | 0.83 |
| Bathing | 3.07 (1.14) | 2.72 (1.22) | –1.38 | .15 | 0.86 |
| Toilet use | 7.13 (2.13) | 6.28 (1.84) | –1.72 | .07 | 0.81 |
| Bowels | 8.63 (1.50) | 9.03 (0.99) | –1.53 | .12 | 0.88 |
| Bladder | 9.04 (1.66) | 9.45 (0.89) | –2.06 | .04 | 0.83 |
| Transfer | 9.50 (2.63) | 9.85 (2.78) | –1.46 | .13 | 0.82 |
| Mobility | 8.75 (2.90) | 8.25 (2.99) | –0.79 | .42 | 0.85 |
| Stairs | 6.67 (2.05) | 6.28 (1.83) | –0.58 | .56 | 0.81 |
| **3-month follow-up** | | | | | |
| Feeding | 8.93 (1.59) | 8.63 (1.38) | –1.77 | .08 | 0.85 |
| Grooming | 4.28 (0.80) | 3.93 (0.58) | –2.85 | .04 | 0.65 |
| Dressing | 7.33 (2.28) | 6.68 (1.81) | –1.86 | .06 | 0.81 |
| Bathing | 3.30 (0.99) | 2.80 (1.09) | –1.93 | .05 | 0.82 |
| Toilet use | 7.25 (1.94) | 6.78 (1.87) | –1.25 | .21 | 0.76 |
| Bowels | 9.23 (1.21) | 9.43 (0.91) | –0.70 | .48 | 0.81 |
| Bladder | 9.03 (1.59) | 9.50 (0.87) | –1.29 | .20 | 0.83 |
| Transfer | 10.38 (2.74) | 10.47 (2.70) | –0.32 | .75 | 0.82 |
| Mobility | 10.32 (3.13) | 9.45 (2.99) | –1.49 | .14 | 0.66 |
| Stairs | 7.05 (2.05) | 6.48 (2.33) | –1.33 | .18 | 0.80 |

## Comparison of Two Mobile Health Follow-Up Assessments With Home Visit Assessments

### *Selecting Candidates for Home Visit Assessments in Two Groups*

We adopted stratified sampling in each group to select subgroups of participants for home visit assessments. We used the discharge MBI scores of 52 participants in the videoconference follow-up group and 57 participants in the telephone follow-up group to classify their functional independence levels into 4 videoconference follow-up subgroups and 4 telephone follow-up subgroups, respectively. The number of participants in the videoconference follow-up and telephone follow-up subgroups is shown in Table 3. We selected half of the participants in each of the 8 subgroups for home visit assessments.

**Table 3.** Distribution of discharge Modified Barthel Index scores in the videoconference follow-up and telephone follow-up subgroups.

| MBI[a] scores | Video follow-up, n (%) | Telephone follow-up, n (%) |
|---|---|---|
| Complete dependence (MBI<40) | 3 (5.8) | 4 (7) |
| Dependence (MBI 40-59) | 17 (32.7) | 16 (28) |
| Mild dependence (MBI 60-99) | 30 (57.7) | 36 (63.1) |
| Independence (MBI 100) | 2 (3.8) | 1 (1.9) |

[a]MBI: Modified Barthel Index.

### *Comparison of Modified Barthel Index Scores Between Videoconference Follow-Up and Home Visit*

Table 4 shows the MBI scores for videoconference follow-up and face-to-face, home visit assessments at 2-week follow-up. MBI scores collected by videoconference were similar to those collected by the face-to-face, home visit assessment, except that a significant difference was found in the feeding task. We also found that ICC values for all 10 tasks were above 0.8, indicating acceptable reliability between the 2 assessments at 2-week follow-up. Table 4 also shows the MBI scores for videoconference follow-up and home visit assessments at

3-month follow-up. Similar comparison results were found between the 2 assessments at 3 months, except that a significant difference was found in the transfer task. ICC values for all but the transfer and stair-climbing tasks exceeded 0.8, indicating acceptable reliability between videoconference and home visit assessments at 3 months.

**Table 4.** A comparison of MBI scores evaluated in videoconference and home visit assessments at 2-week follow-up (n=26) and 3-month follow-up (n=25).

| Variables | Videoconference, mean (SD) | Home visit, mean (SD) | *Z* test | *P* value | Intraclass correlation coefficient |
|---|---|---|---|---|---|
| **2-week follow-up** | | | | | |
| Feeding | 8.04 (1.75) | 7.23 (1.95) | −2.60 | <.001 | 0.87 |
| Grooming | 3.30 (0.88) | 3.50 (0.76) | −1.89 | .06 | 0.90 |
| Dressing | 6.35 (2.21) | 6.58 (2.21) | −1.00 | .32 | 0.92 |
| Bathing | 2.46 (1.27) | 2.54 (1.24) | 0.49 | .62 | 0.87 |
| Toilet use | 6.23 (2.02) | 5.77 (2.08) | −1.63 | .10 | 0.87 |
| Bowels | 9.30 (0.97) | 9.23 (0.99) | −0.58 | .56 | 0.86 |
| Bladder | 9.46 (0.90) | 9.36 (0.96) | −0.09 | .87 | 0.89 |
| Transfer | 9.27 (2.84) | 9.15 (3.51) | −0.72 | .47 | 0.84 |
| Mobility | 8.96 (3.18) | 9.30 (2.99) | 1.0 | .32 | 0.87 |
| Stairs | 6.42 (1.96) | 5.81 (1.83) | −1.83 | .07 | 0.82 |
| **3-month follow-up** | | | | | |
| Feeding | 8.54 (1.42) | 8.12 (1.80) | –1.73 | .08 | 0.83 |
| Grooming | 4.04 (0.53) | 4.12 (0.65) | –1.00 | .32 | 0.88 |
| Dressing | 6.77 (1.75) | 7.04 (1.56) | –1.51 | .13 | 0.92 |
| Bathing | 2.77 (1.10) | 3.00 (1.13) | –1.35 | .18 | 0.82 |
| Toilet use | 6.77 (1.95) | 7.00 (1.90) | 0.81 | .42 | 0.80 |
| Bowels | 9.31 (0.97) | 9.38 (0.94) | –0.58 | .56 | 0.89 |
| Bladder | 9.46 (0.90) | 9.54 (0.86) | –1.00 | .32 | 0.95 |
| Transfer | 10.77 (2.80) | 9.50 (3.31) | –2.41 | .02 | 0.75 |
| Mobility | 9.81 (2.23) | 9.88 (2.76) | 0.14 | .89 | 0.85 |
| Stairs | 6.88 (2.10) | 6.35 (1.65) | –1.48 | .14 | 0.76 |

## Comparison of Modified Barthel Index Scores Between Telephone Follow-Up and Home Visit

Table 5 shows the MBI scores for telephone follow-up and face-to-face, home visit assessments at 2-week follow-up. A comparison of these assessments found that almost all MBI scores collected by telephone administration were statistically higher than those collected by the home visit assessment, with the exception of bowel and bladder management tasks, indicating that the telephone administration method may have overestimated the functional status of study participants for most tasks. ICC values indicate that inadequate reliability was found between the 2 assessment methods at 2 weeks; 8 out of 10 tasks had ICC values less than 0.8.

Table 5 also shows the MBI scores for telephone follow-up and home visit assessments at 3-month follow-up. In general, MBI scores for telephone follow-up were slightly higher than those for home visit assessment, but the only significant differences were found for 4 tasks: feeding, grooming, bathing, and stair climbing. Eight tasks had ICC values less than 0.8, indicating that inadequate reliability was found between the 2 assessment methods at 3 months. Compared with the results of the 2-week follow-up, ICC values showed a general downward trend at the 3-month follow-up.

**Table 5.** Comparison of Modified Barthel Index scores evaluated in telephone and home visit assessments at 2-week follow-up (n=28) and 3-month follow-up (n=24).

| Variables | Telephone, mean (SD) | Home visit, mean (SD) | *Z* test | *P* value | Intraclass correlation coefficient |
|---|---|---|---|---|---|
| **2-week follow-up** | | | | | |
| Feeding | 8.39 (1.85) | 6.79 (1.99) | −3.22 | <.001 | 0.66 |
| Grooming | 3.68 (0.72) | 3.36 (0.78) | −1.97 | .05 | 0.58 |
| Dressing | 6.82 (2.00) | 5.71 (2.02) | −2.52 | .01 | 0.64 |
| Bathing | 3.07 (1.15) | 2.17 (1.19) | -3.15 | <.001 | 0.65 |
| Toilet use | 7.53 (2.11) | 5.25 (2.19) | −3.63 | <.001 | 0.62 |
| Bowels | 8.54 (1.57) | 8.32 (1.83) | −1.03 | .31 | 0.82 |
| Bladder | 8.86 (1.80) | 8.82 (1.80) | −0.30 | .76 | 0.80 |
| Transfer | 10.60 (2.47) | 8.64 (3.23) | −3.07 | <.001 | 0.68 |
| Mobility | 10.93 (2.97) | 8.36 (3.23) | −3.21 | <.001 | 0.63 |
| Stairs | 7.29 (1.88) | 5.61 (1.79) | −3.31 | <.001 | 0.61 |
| **3-month follow-up** | | | | | |
| Feeding | 9.00 (1.53) | 7.37 (2.10) | −3.04 | <.001 | 0.62 |
| Grooming | 4.46 (0.72) | 4.12 (0.61) | −2.20 | .05 | 0.52 |
| Dressing | 7.58 (2.22) | 6.83 (2.18) | −1.50 | .13 | 0.63 |
| Bathing | 3.50 (0.98) | 2.92 (1.02) | −2.08 | .04 | 0.49 |
| Toilet use | 7.17 (2.01) | 6.58 (1.89) | −1.15 | .25 | 0.58 |
| Bowels | 9.20 (1.28) | 9.04 (1.30) | −0.82 | .41 | 0.82 |
| Bladder | 9.08 (1.53) | 9.13 (1.53) | −0.33 | .74 | 0.81 |
| Transfer | 10.54 (2.57) | 9.79 (3.44) | −1.32 | .19 | 0.66 |
| Mobility | 10.37 (2.93) | 9.42 (3.09) | −1.41 | .16 | 0.62 |
| Stairs | 7.29 (1.83) | 6.54 (1.79) | −1.73 | .08 | 0.49 |

## Feasibility of Using the Videoconference and Telephone Function for Collecting Follow-Up Data

As shown in Figure 1 and Table 6, 8 out of 60 (13%) participants in the videoconference follow-up group dropped out, and 3 out of 60 (5%) participants in the telephone follow-up group dropped out at 2-week follow-up. There was no significant difference in completion rates between the 2 groups ($X^2_1$=1.6; *P*=.21) at 2 weeks. At 3-month follow-up, 3 out of 52 (6%) participants in the videoconference follow-up group dropped out, and 9 out of 57 (16%) participants in the telephone follow-up group dropped out. There was no significant difference in the completion rates between the 2 groups ($X^2_1$=1.86; *P*=.17) at 3 months.

**Table 6.** Completion rates at 2-week and 3-month follow-up assessments.

| Completion rates | Videoconference follow-up | Telephone follow-up | $X^2$ (df) | *P* value |
|---|---|---|---|---|
| Completion (2-week follow-up) | 52 | 57 | 1.60 (1) | .21 |
| Dropout (2-week follow-up) | 8 | 3 | N/A[a] | N/A |
| Completion (3-month follow-up) | 49 | 48 | 1.86 (1) | .17 |
| Dropout (3-month follow-up) | 3 | 9 | N/A | N/A |

[a]N/A: not applicable.

Table 7 shows participant ratings of satisfaction, comfort, and confidence with using either the videoconference or telephone call function for follow-up assessments. At 2-week follow-up, the majority of participants were either very satisfied (22/52, 42%) or satisfied (29/52, 56%) with the videoconference function, and either very satisfied (19/57, 33%) or satisfied (34/57, 60%) with the telephone call function. There was no significant difference in the satisfaction levels between the 2 groups ($X^2_3$=2.5; *P*=.28) at 2 weeks. At 3-month follow-up, participants in the videoconference follow-up group reported higher satisfaction than those in the telephone follow-up group ($X^2_3$=13.9; *P*=.03). Additionally, most participants were either

very comfortable (24/52, 46% at 2 weeks; 26/48, 54% at 3 months) or comfortable (28/52, 54% at 2 weeks; 20/48, 42% at 3 months) with the videoconference function, and most participants were either very comfortable (33/57, 58% at 2 weeks; 21/49, 43% at 3 months) or comfortable (24/57, 42% at 2 weeks; 26/49, 53% at 3 months) with the telephone call function. There were no differences in comfort levels between the 2 groups at 2 weeks ($X^2_3$=1.5; *P*=.22) or 3 months ($X^2_3$=1.3; *P*=.52). Regarding participant confidence using the videoconference or telephone function for collecting functional data, participants in the videoconference follow-up group rated higher confidence than those in the telephone follow-up group at 2 weeks ($X^2_3$=6.6; *P*=.04) and 3 months ($X^2_3$=7.9; *P*=.04).

**Table 7.** Ratings of satisfaction, comfort, and confidence at 2-week and 3-month follow-up assessments.

| Ratings | Videoconference follow-up, n (%) | Telephone follow-up, n (%) | $X^2$ (df) | *P* value |
|---|---|---|---|---|
| **Satisfaction (2-week follow-up)** | | | **2.54 (3)** | **.28** |
| Very satisfied | 22 (42) | 19 (33) | | |
| Satisfied | 29 (56) | 34 (60) | | |
| Not very satisfied | 1 (2) | 4 (7) | | |
| Unsatisfied | 0 (0) | 0 (0) | | |
| **Satisfaction (3-month follow-up)** | | | **13.9 (3)** | **.03** |
| Very satisfied | 30 (61) | 12 (25) | | |
| Satisfied | 17 (35) | 30 (63) | | |
| Not very satisfied | 2 (4) | 4 (8) | | |
| Unsatisfied | 0 (0) | 2 (4) | | |
| **Comfort (2-week follow-up)** | | | **1.50 (3)** | **.22** |
| Very comfortable | 24 (46) | 33 (58) | | |
| Comfortable | 28 (54) | 24 (42) | | |
| Not very comfortable | 0 (0) | 0 (0) | | |
| Uncomfortable | 0 (0) | 0 (0) | | |
| **Comfort (3-month follow-up)** | | | **1.30 (3)** | **.52** |
| Very comfortable | 26 (54) | 21 (43) | | |
| Comfortable | 20 (42) | 26 (53) | | |
| Not very comfortable | 2 (4) | 2 (4) | | |
| Uncomfortable | 0 (0) | 0 (0) | | |
| **Confidence (2-week follow-up)** | | | **6.68 (3)** | **.04** |
| Very confident | 33 (63) | 24 (42) | | |
| Confident | 19 (37) | 30 (53) | | |
| Not very confident | 0 (0) | 3 (5) | | |
| Unconfident | 0 (0) | 0 (0) | | |
| **Confidence (3-month follow-up)** | | | **7.97 (3)** | **.04** |
| Very confident | 31 (63) | 20 (42) | | |
| Confident | 18 (37) | 23 (48) | | |
| Not very confident | 0 (0) | 2 (4) | | |
| Unconfident | 0 (0) | 3 (6) | | |

## Discussion

This is one of the first studies to compare the validity and reliability of the videoconference and telephone functions of an mHealth app for collecting functional status data in stroke patients after rehabilitation hospitalization and to examine the feasibility and acceptability of using these modes of administration for data collection. We examined these questions prospectively in a cohort of patients who were discharged from inpatient stroke rehabilitation by comparing videoconference and telephone follow-up assessments, as well as comparing these 2 modes of administration to the home visit assessment as the gold standard.

### Validity and Reliability of the Mobile Health App for Collecting Functional Status Data

We found that most MBI scores obtained by videoconference administration were slightly lower than those obtained by telephone administration, although these score differences did not achieve statistical significance at either 2-week or 3-month periods. Our findings further revealed that videoconference, but not telephone, administration was as valid and reliable as face-to-face, home visit assessment at both 2-week and 3-month follow-up periods. Home visit assessment is conventionally regarded as one of the best methods for collecting posthospitalization outcome measurement data [37]. Home visit (or home rehabilitation) is recognized as providing greater convenience to patients and families while encouraging therapy to continue to occur within the patient's home; however, it is less cost-effective [38]. One of the reasons for this is that home therapists can often only visit one patient at a time. A previous rehabilitation study indicated that the mean amount of time allotted to perform assessments in home visits is 1 hour and 57 minutes (SD 19 minutes) [39]. Additional time is needed to travel to the patient's home, which can be even more time consuming for patients who live at a greater distance. Furthermore, inadequate manpower and financial concerns restrict the implementation of home visit assessments for all patients after hospitalization. Instead, other studies [40] have recommended home-based telemedicine as a viable option in the delivery of poststroke recovery programs, because telemedicine has shown promising results in improving the overall health of stroke patients and in supporting caregivers while being delivered by therapists from a distance. The use of technologies appears to be a novel potential approach for the therapist's assessment of patient performance in home settings. Our findings concur with this notion that an app-based videoconference can be used to assess the functional performance of stroke patients in home settings. The videoconference function may augment other technology-enabled solutions to provide a means of conducting future clinical trials aimed to evaluate the outcomes of any rehabilitation program implemented in the patient's home [41,42].

Our findings revealed that almost all MBI scores obtained by telephone administration were higher than those obtained via home visit. This overestimation of functional scores is particularly obvious at 2-week follow-up; telephone assessment of functional status using the MBI was less reliable compared with the home visit at 2-week follow-up, but reliability did improve at 3-month follow-up. Previous studies have attained strong agreement between telephone and face-to-face assessments [43,44]. Psychometric differences may be partially attributed to the use of diverse scales in different studies. Our study used the MBI, whereas 2 other studies used either the Functional Independence Measure or Modified Rankin Scales to measure poststroke disability. Interestingly, Pietra et al [45] conducted a similar study to compare the validity and reliability of the Barthel Index (BI) administered by telephone compared with face-to-face assessment in patients after stroke. They indicated that telephone assessment with the BI is reliable in comparison with face-to-face assessment. Several possible reasons could explain these differences. First, the measurement tool used in our study was the MBI, which is more rigorous and provides more detailed ratings compared with the BI. Second, the stroke sample in our study consisted of individuals who were discharged from the rehabilitation hospital, whereas the stroke sample in their study consisted of inpatients in the hospital. The use of telephone interview is greatly contingent upon whether patients are cognizant of their self-function. In our study, discharged patients who were living in their home and community settings were more likely to experience their actual function, as they have more opportunities to interact with real-world contextual barriers. This view is further supported by our study findings, wherein we found greater agreement between 3-month telephone and home visit assessments than 2-week telephone and home visit assessments; individuals at the 3-month follow-up point have had more exposure to their real-world environmental barriers and are better able to estimate their functional status.

### Feasibility of the Mobile Health App for Collecting Functional Status Data

It is noteworthy that our findings indicate that completion rates of both videoconference and telephone assessments were greater than 80% at all follow-up periods. These completion rates are within the acceptable range in clinical studies [46]. Moreover, compared with the telephone assessment, patients reported higher satisfaction with and confidence using the videoconference assessment to measure their functional status. A similar study with the WeChat app for health education also revealed high satisfaction perceived by their participants [47]. All of these results confirm that videoconference assessment of the MBI administered via the WeChat app can serve as an alternative tool to the face-to-face, home visit assessment. Videoconference follow-up provides a surveillance platform for clinicians to objectively assess the task performance of patients in their homes. Patients may also perceive a strong sense of participation, which can improve their psychological condition [48]. Prior research [31] and our results have demonstrated the beneficial effects of the WeChat app as a time-effective, cost-effective, and acceptable communication tool for follow-up data collection. However, implementing routine follow-up measurement via technology involves a number of considerations, including the selection of appropriate patients, settings, timing of assessment, and the optimal mode of administration [49]. Offering different modes of administration, such as video consultation, voice communication, text messaging, or image sharing, may help minimize biased sampling and increase patient participation. Future research may consider adopting a mixed method approach that could help to identify facilitators and barriers to the adoption of mHealth apps for collecting posthospitalization data. In addition, the choice of modalities for monitoring patients depends on the size and structure of the organization. An earlier study [50] found that larger organizations report fewer barriers to using technology-based therapeutic tools, likely due to greater resources. Thus, from a researcher or service provider standpoint, this system-level factor is equally important in determining the best mode of administration for follow-up data collection after discharge from inpatient rehabilitation.

### Study Strengths and Limitations

Our study has a number of strengths. Prior research testing the feasibility and psychometrics of mHealth assessments has primarily adopted the observational design, but ours is one of the first few studies to employ the randomized controlled design to compare these characteristics for 2 different mHealth methods to collect outcome data after discharge from inpatient stroke rehabilitation. Additionally, this study conducted 2 follow-up sessions for both assessment methods to better understand any issues related to long-term compliance. Yet, our study has several limitations. First, patients were recruited from one rehabilitation hospital in a coastal province in eastern China; therefore, results may not generalize to persons living in other regions. Second, this study only included participants who owned a mobile phone and served as their own informant to complete the assessment; patients without a mobile phone and those with severe cognitive or communication impairments may have been excluded from this study. Third, the MBI was the only outcome assessment used in the study. Future research should explore mHealth assessments for measuring other health outcomes in poststroke individuals after discharge from the hospital. Another limitation includes measuring acceptability through 3 self-constructed items (ie, satisfaction, comfort, and confidence), which are limited in their measurement of this construct and do not provide actionable data with which to inform future research. Even though these self-constructed items allow for efficient quantification of acceptability, future research should add a qualitative or mixed method approach to provide additional interpretation and meaning to the quantitative results [51]. Furthermore, this study did not record the amount of time required for performing videoconference and telephone follow-up assessments. Future research should measure the duration of these 2 follow-up assessments and compare them with the time needed for the home visit assessment to provide additional validation evidence. Future research is also needed to conduct a randomized controlled trial comparing these 2 modes of mobile administration to traditional telephone interview method (eg, landline phone service) for stroke patients as a control condition.

### Conclusions

This study found satisfactory feasibility and validity of an app-based videoconference method for collecting functional data after inpatient stroke rehabilitation. High completion and acceptability, as well as adequate validity and reliability of the videoconference follow-up method, may support its clinical application in poststroke home rehabilitation programs and long-term health monitoring after hospitalization.

### Acknowledgments

The contents do not necessarily represent the policy of the funding agencies. The authors certify that all financial and material support for this research and work are clearly identified in the manuscript. The authors thank Megen Devine at Washington University for her editorial assistance with this manuscript. This study was funded by the 12th Five-Year Plan Supporting Project of Ministry of Science and Technology of the People's Republic of China (2013BAI10B01). It was supported by the Collaborative Innovation Center of Rehabilitation Technology. The US National Institutes of Health (K01HD095388) supported a portion of the senior author AWKW's effort for developing this manuscript.

### Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 828 KB - mhealth_v8i5e17219_app1.pdf ]

### References

1. Johnston SC, Mendis S, Mathers CD. Global variation in stroke burden and mortality: estimates from monitoring, surveillance, and modelling. Lancet Neurol 2009 Apr;8(4):345-354. [doi: 10.1016/S1474-4422(09)70023-7] [Medline: 19233730]
2. Wang W, Jiang B, Sun H, Ru X, Sun D, Wang L, NESS-China Investigators. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults. Circulation 2017 Feb 21;135(8):759-771. [doi: 10.1161/CIRCULATIONAHA.116.025250] [Medline: 28052979]
3. Broomfield NM, Quinn TJ, Abdul-Rahim AH, Walters MR, Evans JJ. Depression and anxiety symptoms post-stroke/TIA: prevalence and associations in cross-sectional data from a regional stroke registry. BMC Neurol 2014 Oct 1;14:198 [FREE Full text] [doi: 10.1186/s12883-014-0198-8] [Medline: 25269762]
4. Di Carlo A. Human and economic burden of stroke. Age Ageing 2009 Jan;38(1):4-5. [doi: 10.1093/ageing/afn282] [Medline: 19141505]
5. Ferro JM, Caeiro L, Figueira ML. Neuropsychiatric sequelae of stroke. Nat Rev Neurol 2016 May;12(5):269-280. [doi: 10.1038/nrneurol.2016.46] [Medline: 27063107]
6. Boden-Albala B, Litwak E, Elkind MS, Rundek T, Sacco RL. Social isolation and outcomes post stroke. Neurology 2005 Jun 14;64(11):1888-1892. [doi: 10.1212/01.WNL.0000163510.79351.AF] [Medline: 15955939]

7. Wu S, Wu B, Liu M, Chen Z, Wang W, Anderson CS, China Stroke Study Collaboration. Stroke in China: advances and challenges in epidemiology, prevention, and management. Lancet Neurol 2019 Apr;18(4):394-405. [doi: 10.1016/S1474-4422(18)30500-3] [Medline: 30878104]
8. Centers for Disease ControlPrevention (CDC). Outpatient rehabilitation among stroke survivors--21 States and the District of Columbia, 2005. MMWR Morb Mortal Wkly Rep 2007 May 25;56(20):504-507 [FREE Full text] [Medline: 17522589]
9. Zhang L, Yan T, You L, Li K. Barriers to activity and participation for stroke survivors in rural China. Arch Phys Med Rehabil 2015 Jul;96(7):1222-1228. [doi: 10.1016/j.apmr.2015.01.024] [Medline: 25701640]
10. Koh GC, Yen SC, Tay A, Cheong A, Ng YS, de Silva DA, et al. Singapore Tele-technology Aided Rehabilitation in Stroke (STARS) trial: protocol of a randomized clinical trial on tele-rehabilitation for stroke patients. BMC Neurol 2015 Sep 5;15:161 [FREE Full text] [doi: 10.1186/s12883-015-0420-3] [Medline: 26341358]
11. Levine DA, Galecki AT, Langa KM, Unverzagt FW, Kabeto MU, Giordani B, et al. Trajectory of cognitive decline after incident stroke. J Am Med Assoc 2015 Jul 7;314(1):41-51 [FREE Full text] [doi: 10.1001/jama.2015.6968] [Medline: 26151265]
12. Andersen HE, Schultz-Larsen K, Kreiner S, Forchhammer BH, Eriksen K, Brown A. Can readmission after stroke be prevented? Results of a randomized clinical study: a postdischarge follow-up service for stroke survivors. Stroke 2000 May;31(5):1038-1045. [doi: 10.1161/01.str.31.5.1038] [Medline: 10797163]
13. Wu HM, Tang JL, Lin XP, Lau J, Leung PC, Woo J, et al. Acupuncture for stroke rehabilitation. Cochrane Database Syst Rev 2006 Jul 19(3):CD004131. [doi: 10.1002/14651858.CD004131.pub2] [Medline: 16856031]
14. Kwakkel G, Veerbeek JM, van Wegen EE, Wolf SL. Constraint-induced movement therapy after stroke. Lancet Neurol 2015 Feb;14(2):224-234 [FREE Full text] [doi: 10.1016/S1474-4422(14)70160-7] [Medline: 25772900]
15. McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. Med Care 1994 Jun;32(6):551-567. [doi: 10.1097/00005650-199406000-00002] [Medline: 8189774]
16. Faria MC, Mateus CL, Coelho F, Martins R, Barros H. [Mail questionnaires. A useful strategy for the follow-up of patients with a cerebrovascular stroke?]. Acta Med Port 1997 Jan;10(1):61-65 [FREE Full text] [Medline: 9245178]
17. Janssen PM, Visser NA, Mees SM, Klijn CJ, Algra A, Rinkel GJ. Comparison of telephone and face-to-face assessment of the modified Rankin Scale. Cerebrovasc Dis 2010 Jan;29(2):137-139. [doi: 10.1159/000262309] [Medline: 19955737]
18. Chowdhury M, Birns J, Rudd A, Bhalla A. Telemedicine versus face-to-face evaluation in the delivery of thrombolysis for acute ischaemic stroke: a single centre experience. Postgrad Med J 2012 Mar;88(1037):134-137. [doi: 10.1136/postgradmedj-2011-130060] [Medline: 22267526]
19. Lannin NA, Anderson C, Lim J, Paice K, Price C, Faux S, et al. Telephone follow-up was more expensive but more efficient than postal in a national stroke registry. J Clin Epidemiol 2013 Aug;66(8):896-902. [doi: 10.1016/j.jclinepi.2013.03.005] [Medline: 23810029]
20. Savio K, Pietra GL, Oddone E, Reggiani M, Leone MA. Reliability of the modified Rankin Scale applied by telephone. Neurol Int 2013 Feb 11;5(1):e2 [FREE Full text] [doi: 10.4081/ni.2013.e2] [Medline: 23717781]
21. Revicki DA, Tohen M, Gyulai L, Thompson C, Pike S, Davis-Vogel A, et al. Telephone versus in-person clinical and health status assessment interviews in patients with bipolar disorder. Harv Rev Psychiatry 1997;5(2):75-81. [doi: 10.3109/10673229709034730] [Medline: 9385024]
22. Beaver K, Luker KA. Follow-up in breast cancer clinics: reassuring for patients rather than detecting recurrence. Psychooncology 2005 Feb;14(2):94-101. [doi: 10.1002/pon.824] [Medline: 15386784]
23. Goldstein LB, Lyden P, Mathias SD, Colman SS, Pasta DJ, Albers G, et al. Telephone assessment of functioning and well-being following stroke: is it feasible? J Stroke Cerebrovasc Dis 2002;11(2):80-87. [doi: 10.1053/jscd.2002.126691] [Medline: 17903861]
24. Donoff B, McDonough JE, Riedy CA. Integrating oral and general health care. N Engl J Med 2014 Dec 11;371(24):2247-2249. [doi: 10.1056/NEJMp1410824] [Medline: 25494266]
25. Zeng F, Deng G, Wang Z, Liu L. WeChat: a new clinical teaching tool for problem-based learning. Int J Med Educ 2016 Apr 25;7:119-121 [FREE Full text] [doi: 10.5116/ijme.5708.e5c4] [Medline: 27111920]
26. Rasmussen RS, Østergaard A, Kjær P, Skerris A, Skou C, Christoffersen J, et al. Stroke rehabilitation at home before and after discharge reduced disability and improved quality of life: a randomised controlled trial. Clin Rehabil 2016 Mar;30(3):225-236. [doi: 10.1177/0269215515575165] [Medline: 25758941]
27. Statista. Number of Monthly Active WeChat Users From 2nd Quarter 2011 to 3rd Quarter 2019 (in Millions) URL: https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/ [accessed 2020-02-27]
28. Chan WS, Leung AY. Use of social network sites for communication among health professionals: systematic review. J Med Internet Res 2018 Mar 28;20(3):e117 [FREE Full text] [doi: 10.2196/jmir.8382] [Medline: 29592845]
29. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]

30. Liu J, Zheng X, Chai S, Lei M, Feng Z, Zhang X, et al. Effects of using WeChat-assisted perioperative care instructions for parents of pediatric patients undergoing day surgery for herniorrhaphy. Patient Educ Couns 2018 Aug;101(8):1433-1438. [doi: 10.1016/j.pec.2018.02.010] [Medline: 29499997]
31. Lyu KX, Zhao J, Wang B, Xiong G, Yang W, Liu Q, et al. Smartphone application WeChat for clinical follow-up of discharged patients with head and neck tumors: a randomized controlled trial. Chin Med J (Engl) 2016 Dec 5;129(23):2816-2823 [FREE Full text] [doi: 10.4103/0366-6999.194635] [Medline: 27900995]
32. Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. Stroke 2000 Jun;31(6):1429-1438. [doi: 10.1161/01.str.31.6.1429] [Medline: 10835468]
33. Martinsson L, Eksborg S. Activity Index - a complementary ADL scale to the Barthel Index in the acute stage in patients with severe stroke. Cerebrovasc Dis 2006;22(4):231-239. [doi: 10.1159/000094009] [Medline: 16788295]
34. Hsueh I, Wang C, Sheu C, Hsieh C. Comparison of psychometric properties of three mobility measures for patients with stroke. Stroke 2003 Jul;34(7):1741-1745. [doi: 10.1161/01.STR.0000075295.45185.D4] [Medline: 12775883]
35. Leung SO, Chan CC, Shah S. Development of a Chinese version of the Modified Barthel Index-- validity and reliability. Clin Rehabil 2007 Oct;21(10):912-922. [doi: 10.1177/0269215507077286] [Medline: 17981850]
36. Shrout PE. Measurement reliability and agreement in psychiatry. Stat Methods Med Res 1998 Sep;7(3):301-317. [doi: 10.1177/096228029800700306] [Medline: 9803527]
37. Drummond AE, Whitehead P, Fellows K, Sprigg N, Sampson CJ, Edwards C, et al. Occupational therapy predischarge home visits for patients with a stroke (HOVIS): results of a feasibility randomized controlled trial. Clin Rehabil 2013 May;27(5):387-397. [doi: 10.1177/0269215512462145] [Medline: 23113988]
38. Chen AW, Koh YT, Leong SW, Ng LW, Lee PS, Koh GC. Post community hospital discharge rehabilitation attendance: self-perceived barriers and participation over time. Ann Acad Med Singapore 2014 Mar;43(3):136-144 [FREE Full text] [Medline: 24714707]
39. Einarsson U, Gottberg K, Fredrikson S, Bergendal G, von Koch L, Holmqvist LW. Multiple sclerosis in Stockholm County. A pilot study exploring the feasibility of assessment of impairment, disability and handicap by home visits. Clin Rehabil 2003 May;17(3):294-303. [doi: 10.1191/0269215503cr611oa] [Medline: 12735537]
40. Schwamm LH, Holloway RG, Amarenco P, Audebert HJ, Bakas T, Chumbler NR, American Heart Association Stroke Council, Interdisciplinary Council on Peripheral Vascular Disease. A review of the evidence for the use of telemedicine within stroke systems of care: a scientific statement from the American Heart Association/American Stroke Association. Stroke 2009 Jul;40(7):2616-2634. [doi: 10.1161/STROKEAHA.109.192360] [Medline: 19423852]
41. Dobkin BH. A Rehabilitation-Internet-of-Things in the home to augment motor skills and exercise training. Neurorehabil Neural Repair 2017 Mar;31(3):217-227 [FREE Full text] [doi: 10.1177/1545968316680490] [Medline: 27885161]
42. Dobkin BH, Dorsch A. The promise of mHealth: daily activity monitoring and outcome assessments by wearable sensors. Neurorehabil Neural Repair 2011;25(9):788-798 [FREE Full text] [doi: 10.1177/1545968311425908] [Medline: 21989632]
43. Korner-Bitensky N, Wood-Dauphinee S. Barthel Index information elicited over the telephone. Is it reliable? Am J Phys Med Rehabil 1995;74(1):9-18. [doi: 10.1097/00002060-199501000-00003] [Medline: 7873120]
44. Shinar D, Gross CR, Bronstein KS, Licata-Gehr EE, Eden DT, Cabrera AR, et al. Reliability of the activities of daily living scale and its use in telephone interview. Arch Phys Med Rehabil 1987 Oct;68(10):723-728. [Medline: 3662782]
45. Pietra GL, Savio K, Oddone E, Reggiani M, Monaco F, Leone MA. Validity and reliability of the Barthel index administered by telephone. Stroke 2011 Jul;42(7):2077-2079. [doi: 10.1161/STROKEAHA.111.613521] [Medline: 21527755]
46. Richardson PE. David Sackett and the birth of evidence based medicine: how to practice and teach EBM. Br Med J 2015 Jun 8;350:h3089. [doi: 10.1136/bmj.h3089] [Medline: 26055193]
47. Li W, Han LQ, Guo YJ, Sun J. Using WeChat official accounts to improve malaria health literacy among Chinese expatriates in Niger: an intervention study. Malar J 2016 Nov 24;15(1):567 [FREE Full text] [doi: 10.1186/s12936-016-1621-y] [Medline: 27881122]
48. Cao Y, Lin SH, Zhu D, Xu F, Chen Z, Shen H, et al. WeChat public account use improves clinical control of cough-variant asthma: a randomized controlled trial. Med Sci Monit 2018 Mar 14;24:1524-1532 [FREE Full text] [doi: 10.12659/msm.907284] [Medline: 29536984]
49. Wong AW, Heinemann AW, Miskovic A, Semik P, Snyder TM. Feasibility of computerized adaptive testing for collection of patient-reported outcomes after inpatient rehabilitation. Arch Phys Med Rehabil 2014 May;95(5):882-891. [doi: 10.1016/j.apmr.2013.12.024] [Medline: 24440363]
50. Ramsey A, Lord S, Torrey J, Marsch L, Lardiere M. Paving the way to successful implementation: identifying key barriers to use of technology-based therapeutic tools for behavioral health care. J Behav Health Serv Res 2016 Jan;43(1):54-70 [FREE Full text] [doi: 10.1007/s11414-014-9436-5] [Medline: 25192755]
51. Ramsey AT, Wetherell JL, Depp C, Dixon D, Lenze E. Feasibility and acceptability of smartphone assessment in older adults with cognitive and emotional difficulties. J Technol Hum Serv 2016;34(2):209-223 [FREE Full text] [doi: 10.1080/15228835.2016.1170649] [Medline: 27683018]

## Abbreviations

**ADLs:** activities of daily living
**BI:** Barthel Index
**ICC:** intraclass correlation coefficient
**MBI:** Modified Barthel Index
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 28.11.19; peer-reviewed by Y Guo, L Wang; comments to author 18.12.19; revised version received 08.01.20; accepted 07.02.20; published 13.05.20.*

*Please cite as:*
*Li L, Huang J, Wu J, Jiang C, Chen S, Xie G, Ren J, Tao J, Chan CCH, Chen L, Wong AWK*
*A Mobile Health App for the Collection of Functional Outcomes After Inpatient Stroke Rehabilitation: Pilot Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2020;8(5):e17219*
*URL: https://mhealth.jmir.org/2020/5/e17219*
*doi:10.2196/17219*
*PMID:32401221*

Original Paper

# Human-Centered Design Strategies for Device Selection in mHealth Programs: Development of a Novel Framework and Case Study

Ashley Marie Polhemus[1,2], MSc[‡]; Jan Novák[1,3], MSc; Jose Ferrao[1], PhD; Sara Simblett[4], PhD; Marta Radaelli[5], PhD; Patrick Locatelli[6], PhD; Faith Matcham[4,7], PhD; Maximilian Kerz[4], PhD; Janice Weyer[8], BA, BSc; Patrick Burke[8], BSc; Vincy Huang[9], MSc; Marissa Fallon Dockendorf[10], PhD; Gergely Temesi[1], PhD; Til Wykes[4,7], PhD; Giancarlo Comi[5], PhD; Inez Myin-Germeys[11], PhD; Amos Folarin[4,7], PhD; Richard Dobson[4], PhD; Nikolay V Manyakov[12], PhD; Vaibhav A Narayan[13], PhD; Matthew Hotopf[4,7], PhD

[1]Merck Research Labs Information Technology, Merck Sharpe & Dohme, Prague, Czech Republic

[2]Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland

[3]Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, Czech Republic

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

[5]Neurology Services, San Raffaele Hospital Multiple Sclerosis Centre, Milan, Italy

[6]Department of Engineering and Applied Science, University of Bergamo, Bergamo, Italy

[7]National Institute for Health Research, Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, United Kingdom

[8]Patient Advisory Board, Remote Assessment of Disease and Relapse Research Program, King's College London, London, United Kingdom

[9]Merck Research Labs Information Technology, Merck Sharpe & Dohme, Singapore, Singapore

[10]Pharmacokinetics, Pharmacodynamics, and Drug Metabolism, Merck & Co, Inc, Kenilworth, NJ, United States

[11]Department for Neurosciences, Center for Contextual Psychiatry, Katholieke Universiteit Leuven, Leuven, Belgium

[12]Janssen Pharmaceutica NV, Beerse, Belgium

[13]Research and Development Information Technology, Janssen Research & Development, LLC, Titusville, NJ, United States

[‡]the Remote Assessment of Disease and Relapse - Central Nervous System Consortium (www.radar-cns.org)

**Corresponding Author:**
Ashley Marie Polhemus, MSc
Merck Research Labs Information Technology
Merck Sharpe & Dohme
Svornosti 3321/2
Prague, 150 00
Czech Republic
Phone: 420 44 634 48 15
Email: ashley.polhemus@merck.com

## Abstract

**Background:** Despite the increasing use of remote measurement technologies (RMT) such as wearables or biosensors in health care programs, challenges associated with selecting and implementing these technologies persist. Many health care programs that use RMT rely on commercially available, “off-the-shelf” devices to collect patient data. However, validation of these devices is sparse, the technology landscape is constantly changing, relative benefits between device options are often unclear, and research on patient and health care provider preferences is often lacking.

**Objective:** To address these common challenges, we propose a novel device selection framework extrapolated from human-centered design principles, which are commonly used in de novo digital health product design. We then present a case study in which we used the framework to identify, test, select, and implement off-the-shelf devices for the Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) consortium, a research program using RMT to study central nervous system disease progression.

**Methods:** The RADAR-CNS device selection framework describes a human-centered approach to device selection for mobile health programs. The framework guides study designers through stakeholder engagement, technology landscaping, rapid proof of concept testing, and creative problem solving to develop device selection criteria and a robust implementation strategy. It also describes a method for considering compromises when tensions between stakeholder needs occur.

**Results:** The framework successfully guided device selection for the RADAR-CNS study on relapse in multiple sclerosis. In the initial stage, we engaged a multidisciplinary team of patients, health care professionals, researchers, and technologists to identify our primary device-related goals. We desired regular home-based measurements of gait, balance, fatigue, heart rate, and sleep over the course of the study. However, devices and measurement methods had to be user friendly, secure, and able to produce high quality data. In the second stage, we iteratively refined our strategy and selected devices based on technological and regulatory constraints, user feedback, and research goals. At several points, we used this method to devise compromises that addressed conflicting stakeholder needs. We then implemented a feedback mechanism into the study to gather lessons about devices to improve future versions of the RADAR-CNS program.

**Conclusions:** The RADAR device selection framework provides a structured yet flexible approach to device selection for health care programs and can be used to systematically approach complex decisions that require teams to consider patient experiences alongside scientific priorities and logistical, technical, or regulatory constraints.

***(JMIR Mhealth Uhealth 2020;8(5):e16043)*** doi:10.2196/16043

**KEYWORDS**

human-centric design; design thinking; patient centricity; device selection; technology selection; remote patient monitoring; remote measurement technologies

## Introduction

When used as part of health care programs, remote measurement technologies (RMT) such as wearables or biosensors have the potential to affect clinical decision making, provide novel health insights, and improve the standard of care in a variety of disease areas [1-4]. RMT is a subset of mobile health (mHealth) technologies, which includes “any technology that enables monitoring of a person’s health status through a remote interface, which can then be transmitted to a healthcare provider” for review or as a means of education for the user themselves [5]. Though use of RMT in health care programs has grown in recent years [1,2,6,7], its impact on health outcomes does not always live up to its supposed potential [1,7,8].

Successful utilization of RMT depends on careful consideration of the program’s scientific, technical, and usability requirements. However, many programs employ commercially available, “off-the-shelf” devices that cannot be customized according to these requirements. In such cases, program designers are challenged to select devices from hundreds of options [9] in a marketplace where validation is sparse [1,7,8], product turnover is high [10], and relative benefits between device options are often unclear. Comparative studies show either limited accuracy or low to moderate agreement between similar, widely-used devices for common measurements such as activity levels [11-14], sleep [14-16], heart rate [12,17,18], and energy expenditure [14,16,19]. Few industry-wide data standards have been established [6,9,20], and different devices may define and report measurements in ways that are not directly comparable [13]. Additionally, the experiences of potential users—including patients, caregivers, and health care professionals—affect the use of RMT heavily [21-23], but these insights are often not collected or transformed into technology requirements [24]. Unfortunately, RMT that do not cater to user needs can increase patient, caregiver, and health care provider burden in otherwise promising health care programs [6,25] and may negatively impact enrollment and retention [26].

Those designing health care programs often struggle to navigate device selection due to the technology landscape’s complexity and potential tensions between device selection criteria [4,20,27]. To date, few best practices exist to guide the selection of off-the-shelf devices. The Framework of Specifications to Consider During Mobile Technology Selection developed by the Clinical Trial Transformation Initiative lists factors to consider when selecting RMT, including technical performance, data management, safety, and human factors [28]. However, it does not provide a method to apply or prioritize these factors. The Digital Health Selection Framework by the Institute for Healthcare Improvement [29] describes a computational method for assessing the technology landscape based on high-level selection criteria. However, this framework aims to support the development of health care policy, and the method does not support the identification and ranking of sufficiently detailed requirements for use in individual program designs. Scientific publications provide only high-level commentary on device selection, suggesting that designers consider technical requirements, user experiences, data quality, safety, privacy, regulations, costs, and other factors when choosing technologies [27-30]. Such publications also discuss the need to set detailed objectives [27,31] and gather requirements from a diverse set of stakeholders [24,28,31]. However, to our knowledge, no publication describes systematic methods for gathering, prioritizing, and weighing device selection criteria within the context of the program’s users, environments, and goals.

This is problematic, as device-related factors have the potential to limit the success, reproducibility, or scalability of otherwise promising health care programs. In this study we propose a framework to guide device selection based on human-centered design (HCD) principles. We then demonstrate the use of this framework in a research program using RMT to identify and predict relapses in multiple sclerosis (MS).

## Methods

### Human-Centered Design in Mobile Health

HCD is increasingly used to design novel health care programs and products [4,10,32-37]. HCD is a series of methods through which designers study a product user's needs and environment and then design accordingly [38,39]. Designers engage or "empathize" with potential users then generate ideas, develop prototypes, and test those prototypes with the people for whom they are designing [38,39]. Designers alternate between divergent and convergent thinking, looking broadly to understand context and possible solutions, and then converging onto a final problem statement, approach, or solution [38,40]. Many methods also employ agile or lean principles, which use rapid prototyping, feedback loops, and learning cycles to drive an iterative design and implementation process [38,41]. These methods allow designers to develop a deep understanding of the contextual factors that affect design, making them well-suited to support product design in complex, ambiguous, and rapidly-changing environments. The merits of HCD in health care program design have been discussed at length elsewhere [24,33], though such methods are largely applied to de novo designs, rather than technology selection.

HCD frameworks exist for a variety of mHealth applications, including behavioral intervention design [32], implementation of patient-facing technology in interventional clinical trials [31], mHealth solution development and validation [10,33,42], stakeholder engagement [36], and requirement development [43]. Though these frameworks are inconsistent in their language, they employ a set of common methods to inform the design of digital solutions within the context of the health care system (Textbox 1).

**Textbox 1.** Common human-centered design principles recommended in mobile health solution design.

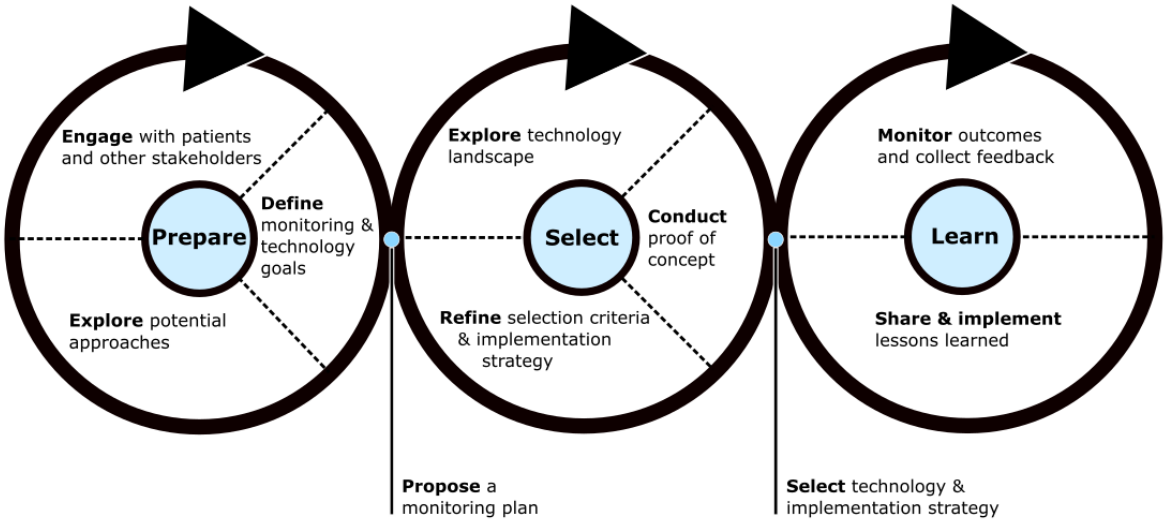
- Assemble a multidisciplinary team [31,43]
- Iterate throughout the design process [10,31-34,36,42,43]
- Begin by conducting stakeholder engagement activities to understand users' needs and environments [31-34,36,42,43]
- Conduct ideation sessions in which a variety of approaches and potential solutions are explored [10,31,32,34,42]
- Enable a variety of stakeholders, including patients, health care professionals, technical experts, and others to participate in the design process [31-34,36,42]
- Prioritize identified requirements and resolve conflicting requirements through further engagement with team members and stakeholders [43]
- Prototype and test with end users prior to scaled implementation [10,31-34,42,43]
- Consider the implementation strategy early and refine it during the design process [31-33]
- Measure the solution's impact and efficacy [10,31,43]
- Share both positive and negative lessons learned with relevant stakeholders to improve current and future designs [31,32]

To our knowledge, no HCD framework addresses the challenges associated with selecting off-the-shelf devices for digital health care interventions. We hypothesized that HCD methods may also be useful for that purpose, because HCD methods address similar design challenges to those posed by device selection. Such challenges include understanding and navigating complicated contextual factors [31,32,34,42,43], engaging with multifunctional stakeholders [36], and prioritizing requirements while addressing diverse stakeholder needs [43].

### RADAR Device Selection Framework

A novel device selection method was developed for the Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) project, a collaborative research program using RMT to study central nervous system disease progression. This framework was developed empirically based on the authors' previous experience with HCD in medical technology design. We hypothesized that HCD methods could help design teams manage the complexity inherent to device selection. Therefore, the three-stage RADAR-CNS device selection framework (Figure 1) was proposed and optimized for the RADAR-CNS program. The framework uses HCD techniques to explore the technology landscape, refine device requirements, develop an implementation strategy, and make informed decisions in parallel with program design and implementation.

**Figure 1.** RADAR device selection framework.




## Stage 1: Prepare

In this stage, the team studies contextual and user-related factors that may affect device use and implementation. The goals of the program, motivations and experiences of patients, involvement of caregivers, and symptoms or sensitivities related to the target disease area will define how user-friendly, discreet, configurable, or multifunctional a device must be. These activities are analogous to the empathize, define, and ideate steps of the design thinking process [44], and similar steps have been proposed in other frameworks [32,33]. In stage 1 we highlight relevant device-related insights that can be collected through HCD methods early in the program planning process.

### *Engage With Patients and Other Stakeholders*

Simblett et al (2018) [22] described five categories of facilitators and barriers that influence patient engagement with RMT: health status, usability, convenience and accessibility, perceived utility, and motivation. During the preparation stage, the device selection team engages with patients and other stakeholders to explore these factors, identify user needs, and draft technology requirements. These activities can be conducted alongside other engagement activities designed to inform program goals or design. Methods for engaging with these and other relevant stakeholders have been proposed, including co-design sessions, focus groups, interviews, workshops, and surveys [44-48]. Integrated patient advisory boards can also guide discussions and decisions throughout the device selection process.

Though published literature on research priorities and usability requirements may provide general insights into patient perspectives in a variety of disease areas [22,24,49], primary research with the program's target population is critical [4,24]. RMT can increase the burden associated with giving and receiving care [4,9,25], which must be minimized to enable sustained program adoption. Direct engagement with potential users provides the nuanced insights that are necessary to minimize burden and increase the chances of program success. Patients may be the primary users of the technology; however, caregivers, health care professionals, and others should also be engaged, as they affect patients' willingness and motivation to engage with RMT [22].

### *Explore Potential Approaches*

The team then explores different approaches for measuring health status. Options should reflect scientific and clinical goals as well as patients' priorities. The team should propose potential *measurement schemes* that list relevant variables or outcomes, surrogate measurements, data streams, required sensors, and desired frequency of measurements. In this stage, it is helpful to use good brainstorming techniques such as those described in IDEO's Design Thinking Bootleg [44] to generate a variety of options and encourage creativity by limiting discussion of potential constraints. The team should define potential program goals, endpoints, and measurement schemes before exploring technology options and implementation strategies [20,27,31]. Delaying discussion of specific technology options forces the team to frame device selection around program and user needs, thereby preventing the design of a program around a familiar but ill-suited technology.

### *Define Measurement and Technology Goals*

Based on the outcomes of the engagement and brainstorming activities, the team should converge on one or more promising measurement schemes and clearly define goals for the RMT. Only once these are defined should the team draft selection criteria. The team should clearly state what compromises they are and are not willing to make, as these choices will drive final device selection. Examples of relevant device selection criteria have been published elsewhere [27-29].

### *Milestone 1: Propose a Monitoring Plan*

By the end of this stage, the team should have developed a robust understanding of stakeholder needs and priorities, a well-defined program goal, one or more potential measurement schemes, and a preliminary understanding of the technology landscape and technology selection criteria. The activities that led to this preliminary plan will provide necessary context to support device selection decisions, especially when no device meets all criteria and concessions must be made. To achieve

this level of clarity, the team may need to conduct multiple iterations of the "Prepare" stage. For example, the team may need to re-engage stakeholders to confirm the acceptability of a measurement scheme and then adjust the scheme in subsequent brainstorming activities.

### Stage 2: Select

In this stage, the team progresses iteratively through a series of activities to identify a suitable device and refine an implementation strategy. With each iteration, the team should identify and answer outstanding questions, refine their thinking, and add detail to their proposed implementation plan. The team should first think broadly before refining the measurement scheme and implementation plan to reflect the program's constraints. This approach allows the team to explore multiple approaches efficiently and to pursue creative options for getting as close to an *ideal* solution as possible.

#### *Explore Technology Landscape*

First, the team performs an initial technology landscape assessment and compiles a list of potentially suitable technologies. Devices should then be systematically excluded from this list based on user feedback and updates to the selection criteria or measurement scheme. When appropriate, additional options should be added to reflect updates to the selection criteria and implementation strategy. A *short list* of candidates should be defined based on the team's selection criteria.

#### *Refine Selection Criteria and Implementation Strategy*

Based on identified technology options and insights from user engagement, the team should begin to define how the technology will be implemented. Factors such as the necessary connections to information technology (IT) systems, device provisioning, training, frequency of device use, compliance monitoring, and data syncing methods should be considered. This strategy may change over time; however, considering these factors early in the selection process will help the team understand potential infrastructure or logistical constraints that could impact device selection. Lack of such strategic planning has been shown to hinder successful implementation of RMT [30].

Off-the-shelf devices may not fit the initial measurement scheme and selection criteria perfectly. Iterative refinement of the selection criteria, measurement scheme, implementation strategy, and technology landscape will help the team explore creative alternatives, make minor concessions, and identify a small group of candidate technologies that meet most criteria.

#### *Conduct Proof of Concept*

Throughout this process, additional questions about candidate devices' characteristics and relative advantages are likely to emerge. In the proof of concept (PoC) phase, the team should conduct targeted tests to answer these questions. PoCs are targeted device assessments that can be conducted quickly prior to implementation in a clinical study that enable rapid learning and decision making during the technology selection process [4,31]. PoCs can test technical characteristics (eg, bench testing for data quality, connectivity, durability), assess user experience in the target population (eg, usability studies), compare candidate devices, or test aspects of a technology's implementation strategy (eg, "dry runs" to test training protocols and technology support systems) [31]. The results of any PoC should be actionable, either in a technology selection decision or to influence refinement of the implementation strategy.

#### *Milestone 2: Select Technology and Implementation Strategy*

By the end of this stage, the team should have narrowed the landscape to a few well-defined technology options, though each is likely to require compromise. To weigh these options, the team should use a systematic method to compare candidate devices and their required compromises. The team should facilitate multifunctional conversations to develop understanding of the required compromises and consensus on a final decision. The team should also finalize an implementation strategy, validating it through PoC testing and additional user feedback as necessary.

### Stage 3: Learn

#### *Monitor Outcomes and Collect Feedback*

The team should also devise mechanisms to collect feedback, experiential data, opportunities for improvement, and opportunities for learning from active programs, and these mechanisms should be included in research protocols if appropriate. Validated questionnaires such as the Post-Study System Usability Questionnaire [50] or the Technology Assessment Model [51] are widely used, and additional quantitative metrics such as device use or help desk engagement rates may also provide insights. Qualitative interviews with patients and health care professionals can identify specific opportunities to improve the implementation strategies, training materials and methods, technologies, or technology support systems.

#### *Share and Implement Lessons Learned*

The design and learning processes should not stop when the program is launched [30]. Quantitative, qualitative, and experiential data collected during all three stages of the framework should be used to continually refine the implementation strategy to ensure efficacy, efficiency, user engagement, ease of use, and clinical utility. In the case of a clinical study where continuous adjustments to the implementation strategy may jeopardize a program's scientific goals, feasibility studies or clinical process evaluations may be used to test and refine the implementation strategy [4,20,52]. Sometimes, devices or technologies selected for an investigational system may not be practical for use in a scaled clinical practice. In this case, appropriate technologies should be selected or designed to fit the system requirements that were collected during investigational implementation. Both positive and negative findings should be shared to inform technology selection decisions in future programs.

## Results

### RADAR-CNS Case Study

RADAR-CNS is a public-private research program leveraging RMT to develop new ways of assessing disease progression in depression, epilepsy, and MS [53]. The RADAR Device

Selection framework was used to select devices for several RADAR-CNS studies; however, only its use in a study on MS disease progression is explored here. In this 2-year study, wearable devices and a custom application collect longitudinal health-related data from people with relapsing-remitting MS. The aim is to develop algorithms that can predict relapse and improve patient care. Details of the study's full protocol are outside the scope of this publication, and only device selection procedures are described here.

## RADAR-CNS: Prepare

A cross-functional team of clinicians, researchers, and technical experts was established, and RADAR-CNS' patient advisory board [54] was also regularly consulted. We worked with people living with MS to understand their perspectives on research priorities, usability requirements, desired device features, and factors influencing sustained engagement with RMT. We conducted a systematic literature review to identify relevant discussion topics [22] and initiated a series of surveys and semistructured focus groups for people living with MS to identify factors affecting engagement with RMT [55]. Participants provided feedback on preferred device features and engagement schemes as well as perspectives on value and privacy. Much of this work has been published previously [55-57]. Participants emphasized the importance in accommodating MS symptoms, making the system easily usable, and enabling users to exert control within the RMT system [55].

We then explored areas of scientific research priority, including cognition, mood, physical activity, sleep, social interactions, speech, and stress. We identified variables that aligned with patient and scientific research priorities, discussed potential measurement schemes, and began to research technological options (eg, data streams, sensors, active tasks, analytical methods). We also began to discuss a variety of technical, user experience, regulatory, and other considerations relevant to the research program. These are described in Multimedia Appendix 1.

### *Milestone 1: Propose a Monitoring Plan*

We prioritized the identified variables based on clinical utility, technological feasibility, alignment with patient priorities, and ethical considerations to select a final measurement scheme for the biosensors (Table 1). Additional clinical, traditional, and mobile data collection methods were also selected, but are outside the scope of this case study. Based on this scheme and patient insights, we defined a preliminary list of required and desired device selection criteria, their relative priorities, and opportunities for compromise. Briefly, the criteria described desired technical capabilities, data quality, user experience, regulatory status, privacy, required investment, and vendor characteristics. Opportunities for compromise included conditions under which multiple devices could be used, acceptable concessions described by patients, and acceptable trade-offs to meet the study budget (eg, willingness to develop bespoke software if device costs are reduced). A summary of these criteria and compromises is available in Multimedia Appendix 1.

**Table 1.** Device-based remote measurement scheme for the RADAR-CNS multiple sclerosis study.

| Factor | Measurement | Measurement Frequency |
|---|---|---|
| Gait | Measured via accelerometer and gyroscope during a 2-Minute Walk Test, tandem walk test, and normal daily activities | Clinical tests[a], home tests[b], free living[c] |
| Balance | Measured via accelerometer placed on the chest during Romberg's Test and normal daily activities | Clinical tests, home tests, free living |
| Fatigue | Measured via heart rate variability and accelerometer during a 2-Minute Walk Test and normal daily activities | Clinical tests, home tests, free living |
| Heart rate and heart rate variability | Measured via one-lead electrocardiogram placed on chest during tests and normal daily activities | Clinical tests, home tests, free living |
| Heart rate and heart rate variability | Measured via photoplethysmography | Daily[d] |
| Sleep | Total sleep time and sleep patterns monitored via actigraphy or other mechanism | Daily |
| Daily Activity | Measured via actigraphy | Daily |

[a]Clinical tests: once every 3 months.

[b]Home tests: once every 3 months.

[c]Free living: one week every 3 months.

[d]Daily: daily over the course of the study.

## RADAR-CNS: Select

We then identified relevant commercially-available consumer and research-grade devices. As no published database contained up-to-date information on available RMT, we conducted an online search and a literature search to identify devices that contained some or all of the sensors in the desired measurement scheme. This search yielded over 100 devices of various embodiments. Devices were systematically excluded through an iterative review process with clinical, analytical, and technical experts, during which potential technologies, priorities, and protocol adjustments were discussed. No single technology fulfilled all selection criteria; however, several devices that fulfilled *most* criteria were selected for further consideration

either as stand-alone devices or for use in conjunction with other devices. These included the Fitbit Charge 2 (Fitbit, Inc., San Francisco, CA), the Withings Steel HR (Withings, Issy-les-Moulineaux, France), the Actigraph Link (ActiGraph LLC, Pensacola, FL), the Suunto Movesense sensor (Suunto Oy, Vantaa, Finland), the eMotion Faros 180 (Biomation, Ottawa, ON, Canada), and the MetaMotion R (MBIENTLAB Inc, San Francisco, CA).

### *Proof of Concept Testing*

Questions regarding usability, data quality, and technical characteristics of the devices arose, prompting appropriate PoC testing of usability, technical features, and training procedures. This section describes two examples of these PoC tests and their impacts on technology selection.

#### Example: User Experience Proof of Concept

Sustained patient engagement with the devices was critical to the study's success, because participants could be enrolled for up to 2 years. The patient advisory board participated in a workshop to provide feedback on candidate devices. Board members, including two members living with MS (authors JW and PB), interacted with each device and provided feedback on user-friendliness, technology preferences, potential impacts of MS symptoms on use, and suggestions for the implementation strategy. This feedback provided us with important context for prioritizing desired device characteristics. The board preferred adhesive patches over chest straps to affix chest-based devices and wrist-based wearables with a subtle or mainstream appearance. They also noted that any goals or feedback shown by the devices, such as daily activity counts, should be customizable. They voiced concern that displaying unrealistic goals could negatively impact participants' motivation to engage with RMT or participate in the study, as people living with MS will almost certainly observe a decline in function over time.

#### Example: Technical Proof of Concept

Following a brainstorming session, the team decided to explore the option of sourcing sensors from an original equipment manufacturer. These devices would be less expensive and more customizable but required additional validation and configuration compared to other options. For commercial reasons, the identities of these devices are not shared. Data were collected from two devices to understand data structure, battery life, reliability of the Bluetooth connection, potential for data loss, data transfer requirements (eg, time, file size, memory availability), and device durability. The devices' published specifications met the requirements; however, the testing demonstrated that neither device met study requirements. The first device's data files were too large to sync more than a few hours of data over a Bluetooth connection, but the study required devices to sync data over Bluetooth outside the clinic. The second device did not meet battery life or data quality requirements in the desired configuration. We tested other candidate devices similarly to address the risks identified by the advisory board and the study teams.

In response to this PoC, we adjusted our technology landscape to include more expensive devices since the tested devices were the only two to meet original budget requirements. To accommodate this change, we also adjusted the implementation strategy to include logistics associated with device returns and reprovisions, thereby reducing the number of required devices and reducing the device cost per patient. This PoC did not yield positive results, but it allowed the team to make data-informed decisions on device candidates without compromising timelines or posing risks to the study.

### *Milestone 2: Select Technology and Implementation Strategy*

Ultimately, we selected 2 devices to conduct all desired measurements. The eMotion Faros 180 was selected to monitor cardiac activity, gait, and balance during home-based active tasks and normal daily activities. The Fitbit Charge 2 was selected to monitor daily activity and sleep based on its superior user experience and battery life, as well as the precedence of Fitbit devices in MS programs [58-60], despite its inability to provide raw accelerometer data. Since no device containing an electrocardiogram, accelerometer, and gyroscope met the necessary criteria, data from the gyroscope sensor in participants' cell phones were collected to identify turns during the 2-Minute Walk Test. A discussion guide used by the team to facilitate the final selection of the wrist-based device is included in Multimedia Appendix 2.

#### RADAR-CNS: Learn

The RADAR-CNS study is ongoing at the time of publication. Surveys and interviews with participants are being conducted periodically throughout the study and device use rates will be monitored as the study progresses. Feedback will also be collected from investigators who conducted the studies. Insights gained through these interactions will be published at the end of the study and will be used to identify improvements to the measurement scheme, device selection, and implementation strategy before the system is available for use in clinical practice.

## Discussion

The RADAR-CNS Device Selection Framework provides methods to assess, prioritize, and adapt device selection criteria for health care programs according to stakeholder needs. The framework is presented linearly, but it is intended to be flexible so teams can move forward, backward, or repeat steps as needed to support device selection. In the RADAR-CNS study, we conducted several iterations of the Prepare and Select stages as our thinking evolved during the study design. These iterations enabled dialogue between the technical and clinical experts on the project, allowing us to establish common ground between stakeholders and ensure consensus on the final decision. We found that our success depended on the engagement of a multifunctional team during each stage of the framework, including investigators, IT specialists, data analysts, patients, health care professionals, and others. Each brought unique perspectives and needs to the process, and each ultimately made compromises to agree on a single technology and implementation strategy. To ensure alignment and mutual understanding between these stakeholders, it was important that members of the device selection team were skilled in

"translating" clinical and technical requirements and their contexts for team members of diverse backgrounds.

Navigating complex stakeholder needs is one of the strengths of HCD, especially when program success is dependent on the willingness of people to continually engage with a technology. As its name suggests, HCD starts by asking designers to understand the people who will be using the technology [38,40,44]. It then enables designers to simultaneously explore program contexts and constraints, identifying connections and priorities between human and nonhuman factors [38,44]. In a systematic review of systematic reviews, Ross et al (2016) [30] found that early engagement with relevant stakeholders such as patients, clinicians, and others was important for successful mHealth implementation, and most frameworks for digital health care solution design echo that sentiment [33]. However, Altman et al (2018) [24] found that user engagement activities were frequently not conducted in such programs. Limited stakeholder centricity during program design and technology selection may ultimately threaten the program's success. Poor user experiences caused by increased burdens [4,26], technical issues [22], lack of accommodations for health status [22], impersonal experiences [26], slowness [22,26], and poor or unclear interface designs [22] may cause patients to stop using the technology, or worse, drop out of the program. Altman et al [24] suggested that, by addressing user needs, HCD methods such as design thinking could increase uptake, adherence, and impact of health care programs that use RMT.

Here, we use HCD methods not to create new designs, but to identify which existing designs are best suited to a particular program. In the RADAR-CNS program, we used HCD methods to identify and prioritize a vast number of often conflicting needs and constraints, not only from patients but also from other "users" of the program: the clinicians caring for patients, the researchers studying diseases, and the technologists developing new monitoring tools. Many common HCD strategies such as empathizing with users, brainstorming, and iterative designing are present in this framework, making it compatible with other HCD approaches to program design or validation.

Though the RADAR Device Selection framework was implemented successfully in an observational research program, its validity in other settings, such as clinical trials of investigational therapies or interventional mHealth program design, must be established in future work. Examples of successful implementation of human-centered methods in health care and academic environments exist; however, their use is not yet routine. Such methods require a mindset shift, new skills, and adoption of additional study planning activities, with more time spent initially on stakeholder engagement [24].

Though selecting off-the-shelf devices for health care programs is often difficult, few best practices exist to guide program designers. To address this gap, we developed and successfully implemented the RADAR device selection framework, which incorporates HCD strategies into a three-stage approach for systematically identifying selection criteria, testing and selecting devices, and monitoring device-related outcomes. To improve RMT implementation in future programs, the methods used and lessons learned during device selection should be more routinely shared.

## Acknowledgments

This paper was written in support of the RADAR-CNS program, a collaborative research effort focusing on the development of RMT for monitoring central nervous system disease progression. We acknowledge all partners in the RADAR-CNS consortium (www.radar-cns.org) for their input and support of this work. The RADAR-CNS project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (www.imi.europa.eu) under grant agreement no 115902. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA; www.imi.europa.eu). This communication reflects the views of the RADAR-CNS consortium, and neither the Innovative Medicines Initiative nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. MH and TW are National Institute of Health Research senior investigators.

## Conflicts of Interest

AP, JN, JF, VH, MD, and GT are employees of Merck Sharp & Dohme Corp, a subsidiary of Merck & Co, Inc, Kenilworth, NJ, USA and may hold stock or stock options in the company. VN and NVM are employed by Janssen Research & Development, LLC and may hold stock or stock options in the company. The remaining authors declare no conflict of interest.

Multimedia Appendix 1
RADAR-CNS device selection considerations and selection criteria.
[DOCX File , 25 KB - mhealth_v8i5e16043_app1.docx ]

Multimedia Appendix 2
RADAR-CNS Multiple Sclerosis Study Device Selection Discussion Guide.
[PPTX File , 443 KB - mhealth_v8i5e16043_app2.pptx ]

## References


1. Noah B, Keller MS, Mosadeghi S, Stein L, Johl S, Delshad S, et al. Impact of remote patient monitoring on clinical outcomes: an updated meta-analysis of randomized controlled trials. NPJ Digit Med 2018;1:20172 [FREE Full text] [doi: 10.1038/s41746-017-0002-4] [Medline: 31304346]
2. Dorsey ER, Topol EJ. State of Telehealth. N Engl J Med 2016 Jul 14;375(2):154-161. [doi: 10.1056/NEJMra1601705] [Medline: 27410924]
3. Marziniak M, Brichetto G, Feys P, Meyding-Lamadé U, Vernon K, Meuth SG. The use of digital and remote communication technologies as a tool for multiple sclerosis management: narrative review. JMIR Rehabil Assist Technol 2018 Apr 24;5(1):e5 [FREE Full text] [doi: 10.2196/rehab.7805] [Medline: 29691208]
4. Polhemus AM, Kadhim H, Barnes S, Zebrowski SE, Simmonds A, Masand SN, et al. Accelerating adoption of patient-facing technologies in clinical trials: a pharmaceutical industry perspective on opportunities and challenges. Ther Innov Regul Sci 2019 Jan;53(1):8-24. [doi: 10.1177/2168479018801566] [Medline: 30373404]
5. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
6. Perry B, Herrington W, Goldsack JC, Grandinetti CA, Vasisht KP, Landray MJ, et al. Use of mobile devices to measure outcomes in clinical research, 2010-2016: a systematic literature review. Digit Biomark 2018;2(1):11-30 [FREE Full text] [doi: 10.1159/000486347] [Medline: 29938250]
7. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
8. 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]
9. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mHealth software applications and wearables to support physical activity assessment, counseling and interventions for cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]
10. Wilson K, Bell C, Wilson L, Witteman H. Agile research to complement agile development: a proposal for an mHealth research lifecycle. NPJ Digit Med 2018;1:46 [FREE Full text] [doi: 10.1038/s41746-018-0053-1] [Medline: 31304326]
11. Chu AHY, Ng SHX, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
12. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
13. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
14. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12(1):42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
15. Ko PT, Kientz JA, Choe EK, Kay M, Landis CA, Watson NF. Consumer sleep technologies: a review of the landscape. J Clin Sleep Med 2015 Dec 15;11(12):1455-1461 [FREE Full text] [doi: 10.5664/jcsm.5288] [Medline: 26156958]
16. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
17. Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, et al. Accuracy of wrist-worn heart rate monitors. JAMA Cardiol 2017 Jan 01;2(1):104-106. [doi: 10.1001/jamacardio.2016.3340] [Medline: 27732703]
18. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
19. Murakami H, Kawakami R, Nakae S, Nakata Y, Ishikawa-Takata K, Tanaka S, et al. Accuracy of wearable devices for estimating total energy expenditure: comparison with metabolic chamber and doubly labeled water method. JAMA Intern Med 2016 May 01;176(5):702-703. [doi: 10.1001/jamainternmed.2016.0152] [Medline: 26999758]
20. Clinical Trials Transformation Initiative. 2018. CTTI Recommendations: Advancing the Use of Mobile Technologies for Data Capture & Improved Clinical Trials URL: https://www.ctti-clinicaltrials.org/files/mobile-devices-recommendations.pdf [accessed 2019-05-30]
21. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

22. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
23. 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]
24. Altman M, Huang TT, Breland JY. Design thinking in health care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
25. Donnelly S, Reginatto B, Kearns O, Mc Carthy M, Byrom B, Muehlhausen W, et al. The burden of a remote trial in a nursing home setting: qualitative study. J Med Internet Res 2018 Jun 19;20(6):e220 [FREE Full text] [doi: 10.2196/jmir.9638] [Medline: 29921563]
26. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
27. Byrom B, Watson C, Doll H, Coons SJ, Eremenco S, Ballinger R, ePRO Consortium. Selection of and evidentiary considerations for wearable devices and their measurements for use in regulatory decision making: recommendations from the ePRO Consortium. Value Health 2018 Jun;21(6):631-639 [FREE Full text] [doi: 10.1016/j.jval.2017.09.012] [Medline: 29909867]
28. Clinical Trials Transformation Initiative. 2018. Framework of Specifications to Consider During Mobile Technology Selection URL: https://www.ctti-clinicaltrials.org/sites/www.ctti-clinicaltrials.org/files/technology-selection-tool.pdf [accessed 2019-05-30]
29. Ostrovsky A, Deen N, Simon A, Mate K. Institute for Healthcare Improvement. Cambridge, MA; 2014. A framework for selecting digital health technology: IHI innovation report URL: http://www.ihi.org/resources/Pages/Publications/AFrameworkforSelectingDigitalHealthTechnology.aspx [accessed 2019-05-30]
30. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
31. TransCelerate BioPharma, Inc. 2018. Patient technology implementation framework URL: https://transceleratebiopharmainc.com/patient-technology-implementation-framework/ [accessed 2019-05-30]
32. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
33. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
34. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
35. Wilhide CC, Peeples MM, Anthony KRC. Evidence-based mHealth chronic disease mobile app intervention design: development of a framework. JMIR Res Protoc 2016 Feb 16;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
36. Sadegh SS, Khakshour Saadat P, Sepehri MM, Assadi V. A framework for m-health service development and success evaluation. Int J Med Inform 2018 Apr;112:123-130. [doi: 10.1016/j.ijmedinf.2018.01.003] [Medline: 29500009]
37. Eslami Andargoli A, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: A systematic review. Int J Med Inform 2017 Jan;97:195-209. [doi: 10.1016/j.ijmedinf.2016.10.008] [Medline: 27919378]
38. Norman D. The Design Of Everyday Things. New York, NY: Basic Books; 2013.
39. Design Kit.: IDEO.org; 2019. What is human-centered design? URL: http://www.designkit.org/human-centered-design [accessed 2019-02-28]
40. Design Council News and Opinion. London, UK: Design Council; 2015. What is the framework for innovation? Design Council's evolved Double Diamond URL: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond [accessed 2019-02-27]
41. Ries E. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation To Create Radically Successful Businesses. New York: Crown Business; 2011.
42. Hevner A. A three cycle view of design science research. Scand J Inf Syst 2007;19(2):7.
43. Van Velsen L, Wentzel J, Van Gemert-Pijnen GJE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
44. Standford DSchool. 2018. Design thinking bootleg URL: https://dschool.stanford.edu/resources/design-thinking-bootleg [accessed 2019-02-28]

45. Woods L, Cummings E, Duff J, Walker K. Partnering in Digital Health Design: Engaging the Multidisciplinary Team in a Needs Analysis. Stud Health Technol Inform 2018;252:176-181. [Medline: 30040702]
46. Woods L, Cummings E, Duff J, Walker K. Design thinking for mHealth application co-design to support heart failure self-management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]
47. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
48. U.S. Food and Drug Administration. 2018. Patient-focused drug development: collecting comprehensive and representative input, guidance for industry, food and drug administration staff, and other stakeholders URL: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm610442.pdf [accessed 2019-05-30]
49. James Lind Alliance. 2018. The PSPs URL: http://www.jla.nihr.ac.uk/priority-setting-partnerships/ [accessed 2019-01-23]
50. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. International Journal of Human-Computer Interaction 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
51. Chin WW, Johnson N, Schwarz A. A fast form approach to measuring technology acceptance and other constructs. MIS Quarterly 2008;32(4):687. [doi: 10.2307/25148867]
52. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
53. RADAR-CNS Consortium. 2019. RADAR-CNS: Remote Assessment of Disease and Relapse URL: https://www.radar-cns.org/ [accessed 2019-01-23]
54. RADAR-CNS Consortium. 2019. Patient Advisory Board | RADAR-CNS URL: https://www.radar-cns.org/patient-engagement/patient-advisory-board [accessed 2019-02-28]
55. Simblett SK, Evans J, Greer B, Curtis H, Matcham F, Radaelli M, RADAR-CNS consortium. Engaging across dimensions of diversity: A cross-national perspective on mHealth tools for managing relapsing remitting and progressive multiple sclerosis. Mult Scler Relat Disord 2019 Jul;32:123-132. [doi: 10.1016/j.msard.2019.04.020] [Medline: 31125754]
56. 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]
57. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
58. DasMahapatra P, Chiauzzi E, Bhalerao R, Rhodes J. Free-living physical activity monitoring in adult US patients with multiple sclerosis using a consumer wearable device. Digit Biomark 2018 Apr 13;2(1):47-63. [doi: 10.1159/000488040]
59. Block VJ, Lizée A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. J Neurol 2017 Feb;264(2):316-326 [FREE Full text] [doi: 10.1007/s00415-016-8334-6] [Medline: 27896433]
60. Sasaki JE, Sandroff B, Bamman M, Motl RW. Motion sensors in multiple sclerosis: Narrative review and update of applications. Expert Rev Med Devices 2017 Nov;14(11):891-900 [FREE Full text] [doi: 10.1080/17434440.2017.1386550] [Medline: 28956457]


## Abbreviations

**CNS:** Central Nervous System
**HCD:** human-centered design
**IT:** information technology
**mHealth:** mobile health
**MS:** multiple sclerosis
**PoC:** proof of concept
**RADAR:** Remote Assessment of Disease and Relapse
**RMT:** remote measurement technologies.

*Edited by G Eysenbach; submitted 02.09.19; peer-reviewed by J Sasaki, M Marcolino; comments to author 09.11.19; revised version received 02.01.20; accepted 24.01.20; published 07.05.20.*

*Please cite as:*
*Polhemus AM, Novák J, Ferrao J, Simblett S, Radaelli M, Locatelli P, Matcham F, Kerz M, Weyer J, Burke P, Huang V, Dockendorf MF, Temesi G, Wykes T, Comi G, Myin-Germeys I, Folarin A, Dobson R, Manyakov NV, Narayan VA, Hotopf M*
*Human-Centered Design Strategies for Device Selection in mHealth Programs: Development of a Novel Framework and Case Study*
*JMIR Mhealth Uhealth 2020;8(5):e16043*
*URL: https://mhealth.jmir.org/2020/5/e16043*
*doi:10.2196/16043*
*PMID:32379055*

Original Paper

# Midwives' Attitudes Toward and Experience With a Tablet Intervention to Promote Safety Behaviors for Pregnant Women Reporting Intimate Partner Violence: Qualitative Study

Lisa Garnweidner-Holme[1], PhD; Lena Henriksen[1], PhD; Eva Marie Flaathen[1], MSc; Tone Klette Bøhler[1], BSc; Mirjam Lukasse[1,2], PhD

[1]Department of Nursing and Health Promotion, Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

[2]Faculty of Health and Social Sciences, University of South-Eastern Norway, Campus Vestfold, Borre, Norway

**Corresponding Author:**
Lisa Garnweidner-Holme, PhD
Department of Nursing and Health Promotion
Faculty of Health Sciences
Oslo Metropolitan University
PB 4 St. Olavs plass
Oslo, 0130
Norway
Phone: 47 48091956
Email: lgarnwei@oslomet.no

## Abstract

**Background:** Violence against women is considered a global health problem, and intimate partner violence (IPV) around the time of childbirth can have severe consequences for mother and child. Prenatal care is considered a window of opportunity to address IPV and ask women about exposure to violence since women are in regular contact with health care providers. Mobile health (mHealth) interventions might overcome the barriers to talking about IPV face-to-face.

**Objective:** Our objective was to explore midwives' attitudes toward a tablet intervention consisting of information about IPV and safety behaviors as well as their experiences with recruiting pregnant women of different ethnic backgrounds in a randomized controlled trial (RCT).

**Methods:** Individual interviews were conducted with 9 midwives who recruited participants for an RCT to test a video to promote safety behaviors delivered on a tablet during prenatal care. Analysis was guided by thematic analysis.

**Results:** Midwives perceived the tablet intervention as an appropriate supplement during prenatal care to provide information about IPV and promote safety behaviors. They participated in the RCT primarily to obtain more knowledge regarding how to communicate about IPV. The intervention was perceived as an anonymous door-opener to talk about IPV and a good solution to ensure that every woman gets the same information. However, the content of the intervention had to be trustworthy and align with the information the midwives provide to women. Given the sensitivity of IPV, midwives outlined the importance of following the intervention with face-to-face communication. Midwives reported technical problems and a high demand on their time as the main challenges to recruiting women. They experienced challenges recruiting women of different ethnic backgrounds due to linguistic barriers and the women's skepticism about scientific research.

**Conclusions:** The tablet intervention might help midwives communicate about IPV. Although the video was considered as an anonymous door-opener to talk about IPV, midwives outlined the importance of following the intervention with face-to-face communication. The scarcity of midwives' time during consultations has to be considered when implementing the intervention. Further research is needed to overcome barriers that limit inclusion of women from different ethnic backgrounds.

**Trial Registration:** ClinicalTrials.gov NCT03397277; https://clinicaltrials.gov/ct2/show/NCT03397277

***(JMIR Mhealth Uhealth 2020;8(5):e16828)*** doi:10.2196/16828

**KEYWORDS**

intimate partner violence; mHealth; attitudes; midwives; prenatal care

## Introduction

Intimate partner violence (IPV) is recognized as a global health problem [1]. According to the definition of the World Health Organization, IPV may include physical aggression, sexual coercion, psychological abuse, or controlling behaviors by current or former partners [1]. Approximately 30% of women worldwide have experienced IPV sometime during their lives [1]. Pregnancy does not protect women from violence. A meta-analysis of IPV during pregnancy that consisted of 92 studies from 23 countries reported an average prevalence of emotional abuse of 28.4%. The prevalences of physical abuse and sexual abuse were 13.8 and 8.0%, respectively [2]. In Norway, studies show the prevalence varies from 1% to 5% [3-5]. Although IPV occurs in all social strata, women with low education levels and women with limited economic resources are at higher risk [6]. Immigrant women are likely to be overrepresented in these groups; hence, they are more prone to be exposed to IPV [6]. The potential impact of IPV on a woman's health prior to pregnancy, during pregnancy, and during the newborn period is well documented and includes depression, miscarriage, perinatal death, preterm birth, and low birth weight [7-9].

Interventions to reduce IPV during pregnancy are urgently needed. Prenatal care is recognized as an ideal window of opportunity to address IPV because this is a time when women are in regular contact with heath care providers [10]. As in many other countries, the Norwegian guidelines for prenatal care strongly recommend that health professionals routinely ask all pregnant women about their experiences with violence [11]. There is some evidence that screening for IPV in prenatal care might increase the identification of violence [12]. However, the evidence regarding how to assess and intervene amid violence during pregnancy and the postpartum period is inconclusive [13-16]. Qualitative studies can be useful in the evaluation of interventions to prevent IPV [17].

Studies among health professionals and women experiencing IPV report several barriers to communicating about IPV face-to-face in clinical settings [18-21]. Mobile health (mHealth) technology such as mobile phones, tablets, and other wireless computing devices may have the potential to overcome some of the barriers regarding face-to-face interventions [22,23]. Health professionals' acceptance of mHealth interventions and willingness to recruit are crucial factors for successful mHealth interventions. Two systematic reviews on health professionals' acceptance found divergent results [24,25]. White et al [24] concluded that health care professionals' acceptance of using mHealth was generally very high. However, health care professionals' acceptance of mHealth interventions during inpatient treatment was low [25]. The authors identified social norms, performance expectancy, and internet integration in the treatment course as the significant predictors for health professionals' acceptance [25]. There is some evidence for the efficacy of Web-based safety decision aids for women experiencing IPV [26]; however, we have not found any published studies about health professionals' acceptance of and experiences with mHealth interventions to prevent IPV during prenatal care.

This study was part of the Safe Pregnancy study, where a tablet intervention to prevent IPV was tested during a randomized controlled trial (RCT) during routine prenatal care at 19 maternal and child health centers (MCHCs) in southeastern Norway. The intervention in the Safe Pregnancy consisted of questions about IPV and a video with information about IPV and safety behaviors [27]. The intervention video lasted for 7 minutes. Only those women who had screened positive for IPV and were randomized to the intervention group saw the video about violence. Women in the control group saw a 7-minute video about a healthy and safe pregnancy in general. All women received a card that included information about referrals for violence. The midwives were blinded to which video the women watched, and although they received information about the content of the videos, the midwives did not watch the videos. The questionnaire and videos were linguistically translated by professional translators into Norwegian, English, Urdu, and Somali and culturally adapted during the development process by women with Norwegian, Pakistani, and Somali backgrounds (as described in the user-involvement study, which has been submitted for publication).

The aim of this qualitative study was to explore midwives' attitudes towards a tablet intervention to prevent IPV and their experiences with recruiting participants with differing ethnic backgrounds to a study that used the tablet intervention.

## Methods

### Interviews

Semistructured individual interviews with 9 midwives who recruited participants for the Safe Pregnancy study were conducted by LG (qualitative researcher) or TK (master's student in midwifery). LG and TK had not met the participating midwives prior to the interviews. Interviews were conducted between June 2019 and September 2019, approximately at the same time when recruitment for the Safe Pregnancy study was finalized at the MCHCs. Interviews were conducted in Norwegian. Of the 9 interviews, 7 interviews took place at the midwives' worksites and 2 at the researchers' worksites. The interviews lasted from 14.00 minutes to 53.38 minutes. The interviews followed a semistructured interview guide (Multimedia Appendix 1) that was developed by the interdisciplinary research group of the Safe Pregnancy study. The interview guide was pilot tested with a midwife who recruited participants for the Safe Pregnancy study. TH conducted the pilot interview in the presence of LG. The pilot test did not lead to any substantial amendments in the final interview guide. The pilot interview was included in the final analysis. The main themes in the interview guide were midwives' motivation to participate in the Safe Pregnancy study, midwives' attitudes toward a tablet intervention to promote safety behaviors, and midwives' experiences with including pregnant women of different ethnic backgrounds.

### Recruitment

Midwives were purposely recruited by LG and TK from the 19 participating MCHCs in the Safe Pregnancy study [27]. The recruitment strategy was to recruit midwives from MCHCs that varied in the cultural backgrounds of their users and number of

recruited women. Recruitment was carried out until a rich set of individual cases was reached.

In total, 9 midwives were asked to participate, and none of the midwives refused to participate. Midwives who helped to recruit women in the Safe Pregnancy study participated in individual teaching sessions on the use of the tablet, how to assess eligibility, and how to recruit women. In addition, midwives had several opportunities for professional development related to IPV, such as attending an international conference about IPV during pregnancy and project workshops with a mix of presentations from various resources from the field of violence against women. The Regional Committee for Medical and Health Research Ethics approved the study (ref.nr: 2017/358), and participants gave their written consent to participate.

### Analysis

Interviews were audiotaped and transcribed by LGH and TK. LGH, LH, TK, and ML read the transcripts. LGH randomly compared some of the transcripts with the audiotapes to ensure the accuracy of the transcription process. The analysis was guided by thematic analysis, according to Braun and Clarke [28], and included the following steps: (1) familiarizing themselves with the data by repeated reading of each informant's transcripts, (2) generating initial codes (words or short phrases in the transcripts) that were relevant for the research questions, (3) organizing the codes into subthemes, (4) arranging the subthemes into overarching themes, and (5) defining and naming the themes. LGH and TK conducted the analysis and discussed potential codes and themes with the other authors. A qualitative software program, HyperRESEARCH 4.0.2 (ResearchWare Inc, Randolph, MA), was used to identify codes and systematize subthemes. In the presentation of the results, midwives are given fictitious names to secure their anonymity.

## Results

### Characteristics of the Midwives

All the midwives worked at the MCHCs, and they routinely asked women about their experiences with violence. Midwives were 42-57 years old. Their length of experience working in prenatal care at MCHCs was 4-18 years. One midwife recruited at 2 MCHCs.

The analysis resulted in 3 themes representing the midwives' attitudes toward a tablet intervention and their experience with recruiting for the Safe Pregnancy study (Textbox 1). The first theme, motivation to participate in the Safe Pregnancy study, represents both the midwives' motivation to participate as well as their general experiences when they asked women about IPV in prenatal care. The second theme, attitudes toward a tablet intervention, describes the midwives' perceived advantages and disadvantages of the use of a tablet during prenatal care to present information about IPV and a safety-promoting video. The third theme, experiences with the recruitment, presents the midwives' general experiences with including women, their suggestions to improve the recruitment process, and their experiences with including women of different ethnic backgrounds.

**Textbox 1.** Themes and subthemes for the midwives' attitudes toward a tablet intervention and experience with recruiting for the Safe Pregnancy study.

**Theme 1: Motivation to Participate in the Safe Pregnancy Study**

- Need for more training and advice for communication about intimate partner violence (IPV) in pregnancy
- Need for more knowledge about IPV in general
- Implementation of a routine enquiry for IPV in pregnancy
- Being part of a network
- Contributing to research

**Theme 2: Attitudes toward a tablet intervention**

- Intervention could act as an anonymous "door-opener" to talk about IPV
- Intervention could make women aware that they experience IPV
- Trust in the content of the tablet intervention
- Importance of face-to-face communication

**Theme 3: Experiences with Recruitment**

- Engaged to recruit
- Time-consuming and disturbing
- Facilitators to recruit
- Experiences including women with different ethnic backgrounds
- Linguistic barriers and skepticism toward scientific research

### Midwives' Motivation to Participate in the Safe Pregnancy Study

For many of the midwives, their motivation to participate in the Safe Pregnancy study was their perceived need for more advice and training about how to ask pregnant women about IPV in order to help them. For example, Anna, who had worked at an MCHC for about 4 years, described her motivation to participate as the following: "It's to get more knowledge and to be more confident to ask."

**Textbox 2.** Example of an exchange between a midwife and moderator.

*You have to find the key for each woman. Every woman has her own key.* [Sara (Midwife)]

*And do you feel that you succeed in finding these keys?* [Moderator]

*Yes.* [Sara]

Midwives who did not perceive it challenging to communicate about IPV were motivated to participate in the study because they were interested in the general knowledge about IPV provided in the workshops, such as the different kinds of violence, adverse health outcomes, and how they could help women.

Midwives often stated that their motivation to participate was the implementation of the national guidelines that strongly recommend midwives ask their clients about IPV, as described by Sandra who had worked long-term in prenatal care: "All of these guidelines came about that we had to ask everybody about violence…and I felt a little bit like 'how do we do that?'"

They were also motivated to become part of a network where they could discuss their experiences and get support from peers. Another motivator to participate was that they wanted to contribute to new scientific knowledge about how to ask about IPV and how to promote safety behaviors during pregnancy.

### Midwives' Attitudes Toward a Tablet Intervention to Inform About Intimate Partner Violence and Promote Safety Behaviors

The majority of the midwives had positive attitudes toward the use of a tablet during prenatal care to provide information about IPV as well as a video to promote safety behaviors, as described by Sandra who worked at an MCHC where many participants were recruited:

> *No, I think that, to answer anonymously on a tablet, that's a nice way to answer nowadays. Because I think that it is difficult to answer this question face-to-face.* [Sandra]

They perceived prenatal care as a good opportunity to use the tablet intervention to promote safety behavior, because as Alexandra stated, "it may help that women are in a safe environment."

Midwives considered a video promoting safety behaviors on a tablet as an appropriate tool for both ethnic Norwegian women and women of different ethnic backgrounds, given that the video is available in various languages. The opportunity for women to answer questions about IPV anonymously and look at safety behaviors on their own was the most frequently mentioned advantage of the tablet intervention. In addition, midwives perceived that the tablet intervention could serve as a door-opener to talk with the women about IPV, as explained by Monica, who had been working for a long time in an urban MCHC with a multiethnic user population:

> *If the video says something like 'your midwife will talk about this and that it is important that she knows, because she has a duty to confidentiality and she can find out how to help you,' (…) That can get people to open up easier. Maybe not the first time one asks, but maybe at a later consultation.* [Monica]

Even though the midwives had not seen the video, they thought that visual information about the different forms of IPV could more easily make women aware that they are victims of violence, compared with face-to-face consultations. Several midwives found it challenging that they had not seen the video and stated that it would be important that they could trust the video content and that the information in the video is in line with the information they provide to women. In this context, midwives said that a potential advantage of the video on a tablet was the possibility to ensure that every woman gets the same information, as stated by Alexandra: "Then you are sure that everyone gets the same [information], maybe."

Many midwives thought that the video could serve as an appropriate supplement to their care and outlined that it would be important that women were followed up face-to-face:

> *I really believe in the face-to-face conversation, with a safe midwife, talking individually…), but one can make people aware about things, in such a video, but it demands a follow-up.* [Ida]

Even though midwives mentioned several advantages of the tablet intervention, some were skeptical since IPV was perceived as a sensitive issue. Alexandra, who had worked at an MCHC for several years, said:

> *But I think that there is no easy solution for this. It's not like that we just can show them a video about it and everything is solved. I don't think so.* [Alexandra]

They said that it would be important to build trust prior to asking women about their experiences with IPV.

This subtheme emerged especially when midwives perceived it challenging to communicate about IPV. These midwives felt that they needed better skills to communicate about IPV to help women and hoped that they could improve their skills through the information in the Safe Pregnancy study's workshops.

Others considered themselves experienced with talking about IPV with pregnant women and did not perceive it as challenging (Textbox 2).

### Midwives' Experiences With Recruiting Women for the Safe Pregnancy Study

In general, midwives appeared to have been very engaged in the recruitment of pregnant women and could even imagine asking women to participate in the Safe Pregnancy study as a door-opener to ask them about IPV. Midwives thought the name of the project, "Safe Pregnancy," might have helped to include women, as it was a general term and not too specific about violence, as described by Sara who also had recruitment experiences from other research projects: "No, I didn't think that it was especially challenging to recruit for a research project about violence, because we should not say that the project was specifically about violence, but that it is about safe pregnancy."

Even though midwives were engaged to recruit, the following statement illustrates that they sometimes felt that they had not succeeded including women who they thought had experienced IPV:

> *Yes, some women are like, you don't really come close to them. I don't know how to explain. But there are women where you think that they should have been included. It's not sure that they have experienced some kind of violence, that's what I don't know, but, yes…* [Alexandra]

Some midwives perceived the recruitment process as disturbing and time-consuming. Midwives explained that they had to deal with many health-related issues in a very short consultation time and having to include women in the Safe pregnancy study disturbed the work. The required time for the intervention was often related to technical problems, usually problems with the internet connection and gaining access to the intervention. Assistance from health secretaries to keep watch for the tablets as well as the availability of separate rooms where participants could be alone with the tablet helped to overcome time-related problems with the recruitment. Some midwives also wanted to have more support and regular follow-up visits by the researchers.

Some midwives experienced challenges including women of different ethnic backgrounds. As expressed by Monica, who recruited at an MCHC with a multicultural population, some midwives found that it was easier to include ethnic Norwegian women: "There have been more nonethnic Norwegian women saying no to participating…"

Midwives reported that well-integrated immigrant women were more willing to participate and mentioned limited language and reading skills as the main barriers to their inclusion. One midwife said that women from Pakistan and Somalia preferred the Norwegian version of the intervention, as they could not read their mother tongue. Midwives thought that the availability of more languages would have made it easier to recruit more women of different ethnic backgrounds.

Another barrier to including women of different ethnic backgrounds was that the midwives often experienced that women of other ethnic backgrounds were skeptical toward research, as described by Anna who has worked at an MCHC with a multicultural user population: "Regarding the questionnaires, for example, women had commented that 'that's not how we talk about these issues in our culture.'"

## Discussion

### Principal Findings

Midwives in this study perceived a tablet intervention as an appropriate supplement to provide information about IPV and to promote safety behaviors during prenatal care. Given the sensitivity of IPV, midwives outlined the importance of following up the intervention with face-to-face communication. Midwives reported technical problems and high time demands as the main challenges to recruiting women. They experienced some challenges including immigrant women due to linguistic barriers and the women's skepticism about scientific research.

### Comparison With Previous Work

The midwives' main motivation to participate in the Safe Pregnancy study was to get more information about IPV and to improve their skills for communicating about IPV to help women. In other studies, health professionals also reported barriers to face-to-face communication about IPV in clinical settings [20,21,29-31]. Discomfort with questioning women about IPV, a fear of offending women, and uncertainty about management after disclosure are the main barriers in face-to-face communication [18,32]. Ongoing mHealth interventions to assess IPV and promote safety behaviors show promising results for overcoming some of the barriers regarding face-to-face communication [22,23,26]. The tablet intervention in the Safe Pregnancy study was perceived as an anonymous door-opener to talk about IPV and provided a good opportunity to increase the women's awareness of violence. A computer tablet intervention was also perceived as a safe and confidential way for abused women to disclose IPV without fear of being judged in a US-based RCT during perinatal home visits [22]. In addition, Glass et al [26] developed and tested a computerized aid to improve the safety decision process in an ethnically diverse sample of abused women in the United States. The aid improved the decision process as demonstrated by reduced decisional conflict after only one use.

As already mentioned, little is known about health professionals' and women's attitudes toward mHealth interventions to promote safety behaviors in order to reduce IPV. Even though the midwives in our study had positive attitudes toward a tablet intervention, they perceived it as an important supplement to face-to-face communication. In line with previous studies, the midwives said that they first had to build trust with the women prior to asking them about IPV [22,31,32]. Women who experienced violence reported a good relationship with the midwife and the trustworthiness of the midwife as a facilitator to talk about IPV [19]. In addition to a trustworthy patient-provider relationship, Bacchus et al [22], who tested an mHealth intervention during perinatal home visits with women experiencing IPV, outlined the importance of considering the clinical context in which interventions for IPV are embedded. The midwives in our study sometimes considered the intervention to be time-consuming and disturbing. Thus, more research is needed about the feasibility of mHealth interventions in prenatal care to promote safety behaviors.

Interventions addressing IPV must consider the cultural and social context where the intervention is implemented [33]. However, communication about IPV seems to be especially challenging between health professionals and women of culturally diverse backgrounds [19]. mHealth offers the opportunity to provide tailored information for different ethnic groups [34]. The Safe Pregnancy study was available in Norwegian, Somali, and Urdu and was culturally adapted during the development process by women with Norwegian, Pakistani, and Somali backgrounds. However, the midwives in our study experienced challenges including immigrant women due to limited language skills and their skepticism toward research projects. This agrees with other studies reporting challenges including women of different ethnic backgrounds [35,36]. Lopez-Class et al [35] identified strategies to recruit immigrant participants. Customizing incentives to specific ethnicities and involvement of local community organizations relevant to immigrants were identified as the most relevant. More research is needed to overcome the barriers related to including immigrant women who have experienced IPV.

### Strengths and Limitations

To our knowledge, this is the first study to investigate health professionals' attitudes toward a tablet intervention during prenatal care to educate women about IPV and promote safety behaviors. The participating midwives represented 10 of the 19 recruitment sites of a large RCT among a multicultural study population; one of the midwives worked at 2 MCHCs. The study included a small sample size, which is typical for qualitative studies [28]. This method was chosen since qualitative studies can contribute to the complex evaluation of mHealth interventions [17]. Only one interview was conducted by LG and TK together, which might have influenced the interviews because of their different educational background and interview style. However, the transcripts and potential themes in the analysis were discussed among all the authors to improve the credibility of the findings [37].

### Conclusions

This study shows that the midwives perceived a video tablet intervention during prenatal care as an appropriate supplement to educate women about IPV and promote safety behaviors. Although the video was considered as an anonymous door-opener to talk about IPV, midwives perceived it important to follow the intervention with face-to-face communication with the women. The scarcity of midwives' time during consultations and technical problems have to be considered when implementing the intervention during prenatal care. Further research is needed to overcome the barriers to including women of different ethnic backgrounds.

## Acknowledgments

We thank the midwives for their participation in this study and their efforts in recruiting participants for the Safe Pregnancy study. This study was funded by the Research Council of Norway Grant nr: 260355. LH was funded by Extrastiftelsen Grant nr: 2016/F076041.

## Authors' Contributions

LGH was involved in the study design, data collection, analysis, and writing of the manuscript. LH was involved in the study design, analysis, and writing of the manuscript. EMF was involved in the study design and writing of the manuscript. TKB was involved in the data collection, analysis, and writing of the manuscript. ML was involved in the study design, analysis, and writing of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary file (interview guide).
[DOCX File , 23 KB - mhealth_v8i5e16828_app1.docx ]

## References

1. Garcia-Moreno C, C. Devries K, Töckl H. Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence. Geneva: World Health Organization; 2013. URL: https://www.who.int/reproductivehealth/publications/violence/9789241564625/en/ [accessed 2020-04-08]
2. James L, Brody D, Hamilton Z. Risk Factors for Domestic Violence During Pregnancy: A Meta-Analytic Review. Violence Vict 2013 Jan 01;28(3):359-380. [doi: 10.1891/0886-6708.vv-d-12-00034]
3. Hjemdal O. E., K. Å spørre om vold ved svangerskapskontroll 2009;Nasjonalt kunnskapssenter om vold og traumatisk stress.
4. Lukasse M, Schei B, Ryding EL. Prevalence and associated factors of fear of childbirth in six European countries. Sexual & Reproductive Healthcare 2014 Oct;5(3):99-106. [doi: 10.1016/j.srhc.2014.06.007]
5. Henriksen L, Schei B, Vangen S, Lukasse M. Sexual violence and mode of delivery: a population-based cohort study. BJOG: Int J Obstet Gy 2014 Jun 18;121(10):1237-1244. [doi: 10.1111/1471-0528.12923]

6. Jewkes R. Intimate partner violence: causes and prevention. The Lancet 2002 Apr;359(9315):1423-1429. [doi: 10.1016/s0140-6736(02)08357-5]
7. Hill A, Pallitto C, McCleary-Sills J, Garcia-Moreno C. A systematic review and meta-analysis of intimate partner violence during pregnancy and selected birth outcomes. International Journal of Gynecology & Obstetrics 2016 Mar 11;133(3):269-276. [doi: 10.1016/j.ijgo.2015.10.023]
8. Alhusen JL, Bullock L, Sharps P, Schminkey D, Comstock E, Campbell J. Intimate Partner Violence During Pregnancy and Adverse Neonatal Outcomes in Low-Income Women. Journal of Women's Health 2014 Nov;23(11):920-926. [doi: 10.1089/jwh.2014.4862]
9. Pastor-Moreno G, Ruiz-Pérez I, Henares-Montiel J, Petrova D. Intimate partner violence during pregnancy and risk of fetal and neonatal death: a meta-analysis with socioeconomic context indicators. American Journal of Obstetrics and Gynecology 2020 Feb;222(2):123-133.e5. [doi: 10.1016/j.ajog.2019.07.045]
10. Devries KM, Kishor S, Johnson H, Stöckl H, Bacchus LJ, Garcia-Moreno C, et al. Intimate partner violence during pregnancy: analysis of prevalence data from 19 countries. Reproductive Health Matters 2010 Nov 24;18(36):158-170. [doi: 10.1016/s0968-8080(10)36533-5]
11. Norwegian DOH, Nasjonal FRFSHAV. National guidlines, ananatal care ? how to uncover violence (in Norwegian). 2014: Oslo 2014.
12. O'Doherty LJ, Taft A, Hegarty K, Ramsay J, Davidson LL, Feder G. Screening women for intimate partner violence in healthcare settings: abridged Cochrane systematic review and meta-analysis. BMJ 2014 May 12;348(may12 1):g2913-g2913. [doi: 10.1136/bmj.g2913]
13. Pallitto C, García-Moreno C, Stöeckl H, Hatcher A, MacPhail C, Mokoatle K, et al. Testing a counselling intervention in antenatal care for women experiencing partner violence: a study protocol for a randomized controlled trial in Johannesburg, South Africa. BMC Health Serv Res 2016 Nov 5;16(1). [doi: 10.1186/s12913-016-1872-x]
14. Stewart DE, Aviles R, Guedes A, Riazantseva E, MacMillan H. Latin American and Caribbean countries' baseline clinical and policy guidelines for responding to intimate partner violence and sexual violence against women. BMC Public Health 2015 Jul 15;15(1). [doi: 10.1186/s12889-015-1994-9]
15. Tiwari A. A randomised controlled trial of empowerment training for Chinese abused pregnant women in Hong Kong. Bjog 2005;112(9):1249-1256. [doi: 10.1111/j.1471-0528.2005.00709.x]
16. Kiely M, El-Mohandes AAE, El-Khorazaty MN, Gantz MG. An Integrated Intervention to Reduce Intimate Partner Violence in Pregnancy. Obstetrics & Gynecology 2010;115(2, Part 1):273-283. [doi: 10.1097/aog.0b013e3181cbd482]
17. Maar MA, Yeates K, Perkins N, Boesch L, Hua-Stewart D, Liu P, et al. A Framework for the Study of Complex mHealth Interventions in Diverse Cultural Settings. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e47. [doi: 10.2196/mhealth.7044]
18. Eustace J, Baird K, Saito AS, Creedy DK. Midwives' experiences of routine enquiry for intimate partner violence in pregnancy. Women and Birth 2016 Dec;29(6):503-510. [doi: 10.1016/j.wombi.2016.04.010]
19. Garnweidner-Holme LM, Lukasse M, Solheim M, Henriksen L. Talking about intimate partner violence in multi-cultural antenatal care: a qualitative study of pregnant women's advice for better communication in South-East Norway. BMC Pregnancy Childbirth 2017 Apr 19;17(1). [doi: 10.1186/s12884-017-1308-6]
20. Henriksen L, Garnweidner-Holme L, Thorsteinsen K, Lukasse M. 'It is a difficult topic' – a qualitative study of midwives´ experiences with routine antenatal enquiry for intimate partner violence. BMC Pregnancy Childbirth 2017 Jun 2;17(1). [doi: 10.1186/s12884-017-1352-2]
21. Sprague S, Madden K, Simunovic N, Godin K, Pham NK, Bhandari M, et al. Barriers to Screening for Intimate Partner Violence. Women & Health 2012 Aug;52(6):587-605. [doi: 10.1080/03630242.2012.690840]
22. Bacchus LJ, Bullock L, Sharps P, Burnett C, Schminkey DL, Buller AM, et al. Infusing Technology Into Perinatal Home Visitation in the United States for Women Experiencing Intimate Partner Violence: Exploring the Interpretive Flexibility of an mHealth Intervention. J Med Internet Res 2016 Nov 17;18(11):e302. [doi: 10.2196/jmir.6251]
23. World HO, mHealth. New horizons for health through mobile technologies: second global survey on eHealth. 2011: Geneva 2011.
24. White A, Thomas DS, Ezeanochie N, Bull S. Health Worker mHealth Utilization. CIN: Computers, Informatics, Nursing 2016;34(5):206-213. [doi: 10.1097/cin.0000000000000231]
25. Hennemann S, Beutel ME, Zwerenz R. Ready for eHealth? Health Professionals' Acceptance and Adoption of eHealth Interventions in Inpatient Routine Care. Journal of Health Communication 2017 Feb 19;22(3):274-284. [doi: 10.1080/10810730.2017.1284286]
26. Glass N, Eden KB, Bloom T, Perrin N. Computerized Aid Improves Safety Decision Process for Survivors of Intimate Partner Violence. J Interpers Violence 2009 Dec 29;25(11):1947-1964. [doi: 10.1177/0886260509354508]
27. Henriksen L, Flaathen EM, Angelshaug J, Garnweidner-Holme L, Småstuen MC, Noll J, et al. The Safe Pregnancy study - promoting safety behaviours in antenatal care among Norwegian, Pakistani and Somali pregnant women: a study protocol for a randomized controlled trial. BMC Public Health 2019 Jun 10;19(1). [doi: 10.1186/s12889-019-6922-y]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

29. Hjemdal O. K. and K. Engnes, Å spørre om vold ved svangerskapskontroll 2009;Nasjonalt kunnskapssenter om vold og traumatisk stress:92.
30. Baird K, Creedy D, Eustace J, Saito A. Australian midwives' understanding and experience of routine enquiry for intimate partner violence during pregnancy. Women and Birth 2015;28:S39-S40. [doi: 10.1016/j.wombi.2015.07.126]
31. Smith R, Wight R, Homer CS. ' Asking the hard questions ': Improving midwifery students' confidence with domestic violence screening in pregnancy. Nurse Education in Practice 2018 Jan;28:27-33. [doi: 10.1016/j.nepr.2017.05.006]
32. Henriksen L, Garnweidner-Holme L, Thorsteinsen K, Lukasse M. 'It is a difficult topic' – a qualitative study of midwives´ experiences with routine antenatal enquiry for intimate partner violence. BMC Pregnancy Childbirth 2017 Jun 2;17(1). [doi: 10.1186/s12884-017-1352-2]
33. Schwab-Reese LM, Renner LM. Screening, management, and treatment of intimate partner violence among women in low-resource settings. Womens Health (Lond Engl) 2018 Apr 06;14:174550651876670. [doi: 10.1177/1745506518766709]
34. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth Research: A Decade of Evolution. J Med Internet Res 2013 May 21;15(5):e95. [doi: 10.2196/jmir.2430]
35. Lopez-Class M, Cubbins L, Loving AM. Considerations of Methodological Approaches in the Recruitment and Retention of Immigrant Participants. J. Racial and Ethnic Health Disparities 2015 Jul 14;3(2):267-280. [doi: 10.1007/s40615-015-0139-2]
36. Garnweidner-Holme L, Hoel Andersen T, Sando MW, Noll J, Lukasse M. Health Care Professionals' Attitudes Toward, and Experiences of Using, a Culture-Sensitive Smartphone App for Women with Gestational Diabetes Mellitus: Qualitative Study. JMIR Mhealth Uhealth 2018 May 14;6(5):e123. [doi: 10.2196/mhealth.9686]
37. Denzin N. K. and Y. S. Lincoln, Strategies of Qualitative Inquiry. 3 ed 2008;Thousand Oaks, California.


## Abbreviations

**IPV:** intimate partner violence.
**MCHC:** mother and child health center.
**mHealth:** mobile health.
**RCT:** randomized controlled trial.


*Edited by G Eysenbach; submitted 29.10.19; peer-reviewed by L Schwab-Reese, M Herdeiro; comments to author 10.03.20; revised version received 13.03.20; accepted 20.03.20; published 20.05.20.*

*Please cite as:*
*Garnweidner-Holme L, Henriksen L, Flaathen EM, Klette Bøhler T, Lukasse M*
*Midwives' Attitudes Toward and Experience With a Tablet Intervention to Promote Safety Behaviors for Pregnant Women Reporting Intimate Partner Violence: Qualitative Study*
*JMIR Mhealth Uhealth 2020;8(5):e16828*
*URL: http://mhealth.jmir.org/2020/5/e16828/*
*doi:10.2196/16828*
*PMID:32432553*

Original Paper

# A Mobile Phone–Based Gait Assessment App for the Elderly: Development and Evaluation

Runting Zhong[1,2], PhD; Pei-Luen Patrick Rau[2], PhD

[1]School of Business, Jiangnan University, Wuxi, China

[2]Department of Industrial Engineering, Tsinghua University, Beijing, China

**Corresponding Author:**
Pei-Luen Patrick Rau, PhD
Department of Industrial Engineering
Tsinghua University
Shunde Building, 5th Floor
Beijing, 100084
China
Phone: 86 62776664
Email: rpl@mail.tsinghua.edu.cn

## Abstract

**Background:** Gait disorders are common among older adults. With an increase in the use of technology among older adults, a mobile phone app provides a solution for older adults to self-monitor their gait quality in daily life.

**Objective:** This study aimed to develop a gait-monitoring mobile phone app (Pocket Gait) and evaluate its acceptability and usability among potential older users.

**Methods:** The app was developed to allow older adults to track their gait quality, including step frequency, acceleration root mean square (RMS), step regularity, step symmetry, and step variability. We recruited a total of 148 community-dwelling older adults aged 60 years and older from two cities in China: Beijing and Chongqing. They walked in three ways (single task, dual task, and fast walking) using a smartphone with the gait-monitoring app installed and completed an acceptability and usability survey after the walk test. User acceptability was measured by a questionnaire including four quantitative measures: perceived ease of use, perceived usefulness, ease of learning, and intention to use. Usability was measured using the System Usability Scale (SUS). Interviews were conducted with participants to collect open-ended feedback questions.

**Results:** Task type had a significant effect on all gait parameters, namely, step frequency, RMS, step variability, step regularity, and step symmetry (all *P* values <.001). Age had a significant effect on step frequency (*P*=.01), and region had a significant effect on step regularity (*P*=.04). The acceptability of the gait-monitoring app was positive among older adults. Participants identified the usability of the system with an overall score of 59.7 (SD 10.7) out of 100. Older adults from Beijing scored significantly higher SUS compared with older adults from Chongqing (*P*<.001). The age of older adults was significantly associated with their SUS score (*P*=.048). Older adults identified improvements such as a larger font size, inclusion of reference values for gait parameters, and inclusion of heart rate and blood pressure monitoring.

**Conclusions:** This mobile phone app is a health management tool for older adults to self-manage their gait quality and prevent adverse outcomes. In the future, it will be important to take factors such as age and region into consideration while designing a mobile phone–based gait assessment app. The feedback of the participants would help to design more elderly-friendly products.

***(JMIR Mhealth Uhealth 2020;8(5):e14453)*** doi:10.2196/14453

**KEYWORDS**

aged; gait; mHealth; telemedicine; falls prevention

## Introduction

### Background

With an increasing aging population, 11.9% of the Chinese population was 65 years and older in 2018 [1]. Approximately 28% to 35% of older adults aged 65 years and older fall each year [2]. Gait assessment is useful for older adults because they are vulnerable to frailty or fall risk. For example, gait speed is a well-known indicator of functional ability [3]; stride-to-stride variability might be a predictor of a future fall [4] or frailty

[5,6]. Moreover, gait characteristics have been found to be associated with cognitive impairment [7]. Early detection of the decline of gait parameters may help older adults adopt timely interventions, such as gait training, to maintain health and improve their quality of life.

Regarding the influence of age on gait parameters, studies seem to reveal conflicting results. According to Menz et al [8], older adults exhibited a more conservative gait pattern compared with younger adults, characterized by reduced velocity, shorter step length, reduced acceleration root mean square (RMS), and increased step timing variability. Koss et al [9] derived multiple gait parameters from an iPod to predict age-related gait changes and found that younger adults had a more variable, less predictable, and more symmetric gait pattern compared with older adults.

Dual tasks (DTs, walking while conducting a secondary task) are usually applied in gait studies to amplify the effects as they require additional cognitive resources and are common for older adults in daily life [10,11]. A demanding task (eg, fast walking [FW] and the addition of a cognitive distraction) might enhance the sensitivity and specificity of frailty prediction and is recommended for frailty assessment using gait analysis [12]. Smith et al [11] found that a cognitive task was more challenging for older adults than a motor task when they were performing the timed up and go (TUG) test. DT gait may result in decreased walking speed [13] and step frequency [13], increased step time variability [10] and stride time variability [10], and double support time variability [10].

Older adults are increasingly using smartphones and mobile apps. Providing health-related apps for gait assessment would help older adults improve their health outcomes and reduce the burden of care. Several studies have tested validity and reliability of gait analysis using a smartphone [14-17]. Nishiguchi et al [18] found that the reliability and validity of a smartphone in measuring step variability, autocorrelation, and acceleration RMS was comparable with an external accelerometer. Manor et al [15] created an iPhone app for assessment of normal and DT walking and found that the app was valid and reliable in measuring stride timing, compared with the gold standard—instrumented GAITRite mat (CIR Systems, PA) [15]. In a previous study, we established reference gait parameters (walking speed, step frequency, RMS, amplitude variability, step variability, step regularity, and step symmetry) of nonfrail and prefrail older adults under single tasks (STs) and DTs [13]. The study found that prefrail older adults showed significantly decreased speed, mediolateral RMS, vertical RMS, anteroposterior RMS, vertical amplitude variability, and vertical step regularity compared with nonfrail older adults [13]. However, there may be a bias about whether older adults will accept using such technologies in their daily life. To overcome such challenges, it is necessary to involve older adults in the evaluation of such apps to improve design and acceptability.

Existing studies examined older adults' acceptance of a health-related app. A study by Liu et al [19] reveals the possibility to predict users' technology acceptance with socioeconomic variables [19]. According to a survey conducted in Hong Kong, 24.10% (995/4129) smartphone or tablet owners had a health app. Tracking physical activity (67.0%, 667/995) and logging health records (43.0%, 428/995) were the most common functions of the health apps. Overall, a younger age, higher education, and higher household income were associated with having health apps. Engaging in moderate physical activity (≥1 day/week, compared with physical inactivity) and having a history of chronic diseases were also associated with having health apps. The study showed a lower prevalence of use of information and communication technologies (ICTs) in respondents with lower education and income in the most developed Chinese city. This could be seen as a confirmation of the *inverse information law*, which suggests that those most in need have less use of services, and hence, receive less benefits from advancements in health-related ICTs [20]. Inspired by this phenomenon, it is necessary to investigate the acceptability among older adults from different socioeconomic positions to contextually inform specific policies to promote the app. In this study, we used two cities Beijing and Chongqing with different socioeconomic levels in China as examples. Beijing is in North China, with an average gross domestic product (GDP) of 128,994 Chinese Yuan (CNY), whereas Chongqing is in West China, with an average GDP of 63,442 CNY [21].

### Objectives

The study had two aims: (1) develop a gait-monitoring mobile phone app (Pocket Gait) and evaluate its acceptability and usability among potential older users, and (2) conduct gait assessment using the app and examine the influence of age group, task type, and region. The main contribution of this study was that we developed low-cost mobile phone apps using an Android smartphone (vivo Z1, Android operating system version 8.1, VIVO Technology Co, China) compared with the gold standard—instrumented GAITRite mat. The app discussed in this paper could assist with daily gait assessment. In addition, the acceptability and usability results could provide design recommendations to promote use of the app among Chinese older adults.

## Methods

### Gait Assessment App Development

#### *Key Design Considerations*

*Pocket Gait* was designed to achieve the goal of monitoring gait quality in daily life. As an accelerometer is commonly embedded in smartphones, it could be used to collect gait data. A previous study illustrated the importance of tracking and giving feedback [22]. Older adults suggested that the data display should enable users to understand the results better. For example, a gait analysis report is required to explain the results with graphs, conclusions, and medical advice [22]. The design of the gait assessment app was based on this idea. The key design requirements are listed as follows:

- Gait test: Users wore a smartphone on the third lumbar spine vertebra (L3) region of the back [23], as shown in Figure 1. When the user was walking, the smartphone would start collecting acceleration data of the three axes.
- Viewing graph: The interface displayed the vertical acceleration pattern to reveal the periodicity of walking. A

researcher could also check if the data were being collected properly.
- Viewing report: The interface displayed the critical gait parameters of straight walking (step frequency, step intensity [RMS], step regularity, step symmetry, and step variability).
- Send: Users sent the gait data to researchers via email. Meanwhile, the raw data were stored locally in the smartphone.

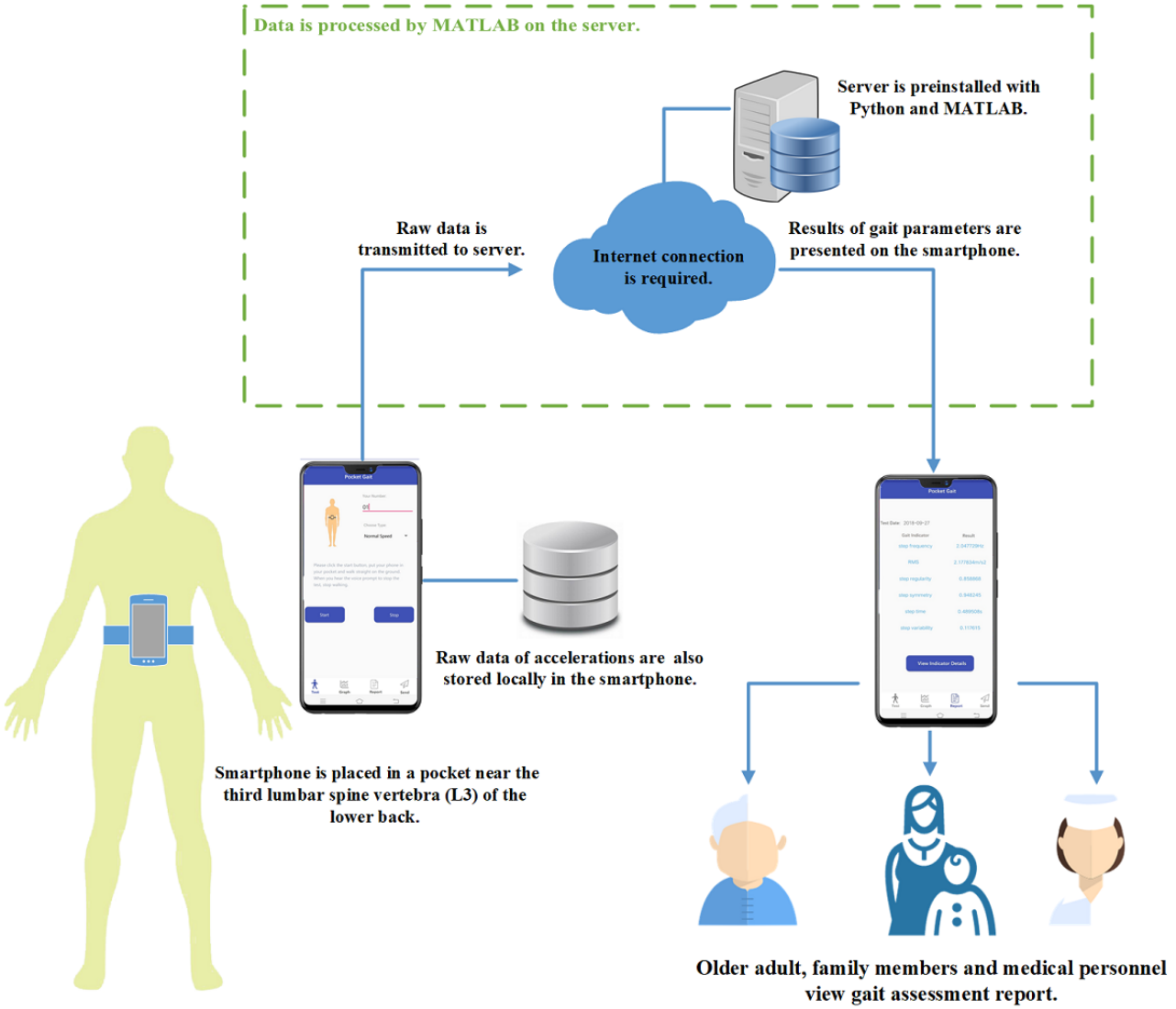
### *System Architecture*

The gait assessment app was developed on an Android smartphone to allow gait data collection and presentation of the results. The system architecture was set up as follows:

- The Android smartphone served as a client, detecting motion (acceleration) when the user was walking and sending these data to the server.
- The server was installed on a computer with Python (Python Software Foundation) and MATLAB (Mathworks, Natick, MA) preinstalled, receiving and processing gait data from the smartphone.
- The smartphone received the gait assessment results and displayed them on the screen.
- Figure 2 illustrates the system architecture of Pocket Gait.

**Figure 1.** The smartphone was placed in a pocket near the third lumbar spine vertebra (L3) of the lower back. The screen of the smartphone faced outward.

**Figure 2.** System architecture of the Pocket Gait app.




### *System Development*

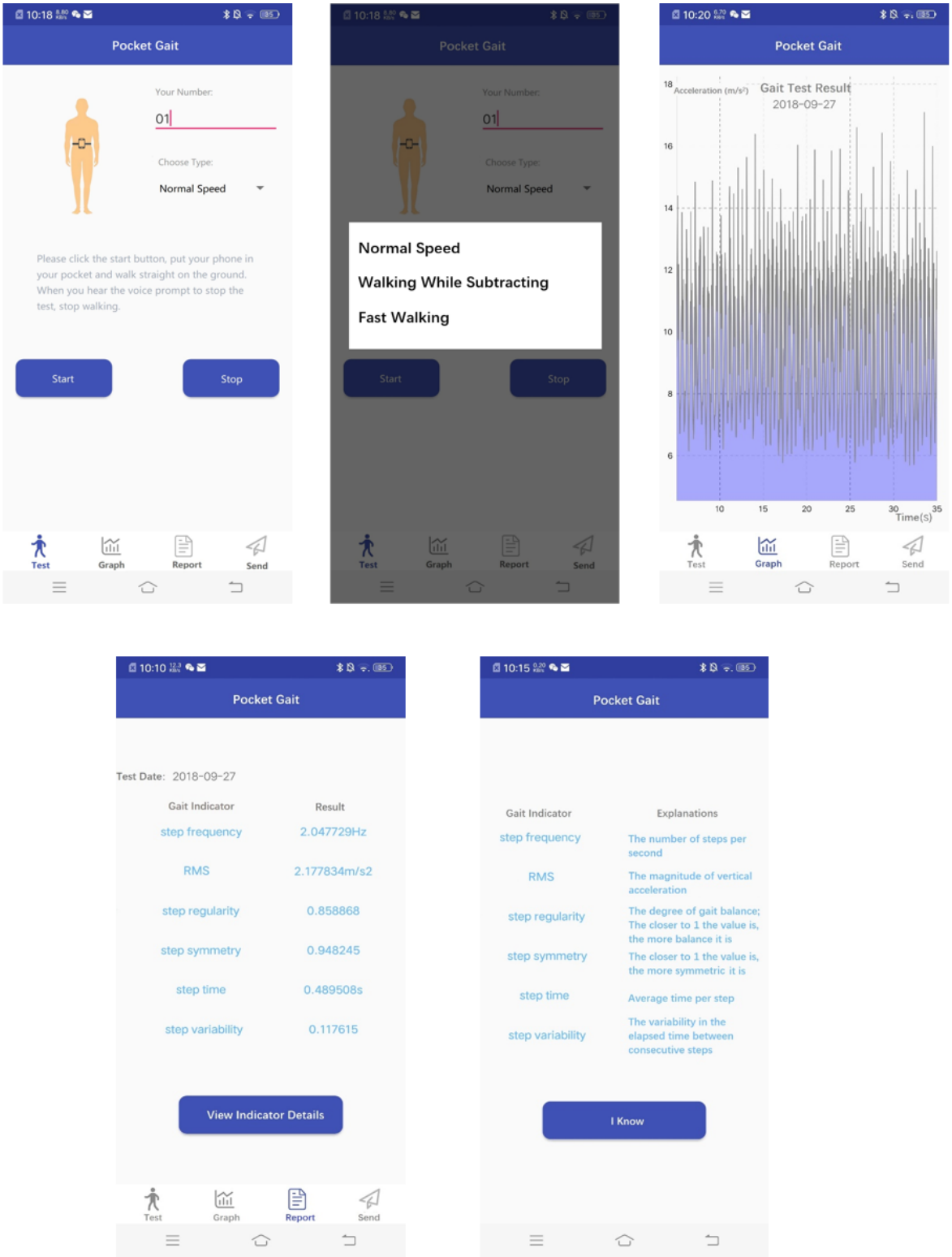
The gait-monitoring app was initially developed with the overall goal to monitor gait quality, including step frequency, acceleration RMS, step regularity, step symmetry, and step variability. The algorithm for the gait parameters was based on a previous study and developed using a self-designed MATLAB program [13]. A detailed description of the algorithm can be found in Multimedia Appendix 1.

The mobile phone app was developed using Android Studio, as Android is also the most popular operating system among Chinese smartphone users, with almost 80% of the share as of July 2017 [24]. The sampling rate of acceleration measurement for the smartphone was set at the highest mode listed in the specifications for an Android smartphone, which is SENSOR_DELAY_FASTEST [25]. The actual sampling rate was around 40 Hz. The initial 5 seconds were not included in data collection to avoid the influence of the acceleration process. When the user pressed the *Start* button, the app would collect the data from the 5th second to 35th second. In other words, the app would collect 30 seconds of the walking data. There were voice instructions about how to walk when the user pressed the *Start* button. In addition, there were voice instructions reminding the user to stop walking at the end of the gait assessment. The app had three walking types: ST, cognitive DT, and FW. For ST, the participant should walk at normal speed. For DT, the participant should walk while serially subtracting 3 from a 3-digit number randomly given by the experimenter, stating the answers out loud. For FW, the participant should walk as fast as possible.

When pressing the *Graph* button in the navigation bar, the interface would display the vertical acceleration pattern to reveal the periodicity of walking.

When pressing the *Report* button in the navigation bar, the collected acceleration data were automatically uploaded to a remote server via Wi-Fi. The gait indicators step frequency, acceleration RMS, step variability, step regularity, and step symmetry were then calculated by the MATLAB engine on the remote server and displayed on a smartphone screen, as shown in Figure 3.

**Figure 3.** Screenshots of the Pocket Gait app: the gait test page, the choosing walking type page, the viewing graph page, the report page, and the explanation page.




## Empirical Data Collection and User Evaluation

To evaluate the proposed prototype, we conducted user evaluations in a corridor over a distance of about 40 meters, collecting evaluations from participants after using the app.

### *Participant Recruitment*

In October 2018, a total of 148 older adults were recruited from universities and nearby communities in Beijing (n=70) and Chongqing (n=78). The older adults were recruited through recruitment flyers, word of mouth, and social media. The inclusion criteria were (1) age ≥60 years, (2) living independently in the community, and (3) being able to walk independently without an assistive device for at least 40 meters. Participants were excluded if they had any musculoskeletal or neurological disease, or painful conditions, that could affect gait. Tsinghua University gave ethical approval for the study.

Each participant was asked to provide written informed consent before participation. Each participant was given 80 CNY after he or she completed the experiment.

### *Study Design*

The experiment used a 2×3×3 mixed design. The 3 independent variables were the region (Beijing or Chongqing), task type (ST, DT, or FW) and age group (60–69, 70–79, or 80–89 years). The between-subject variables were region and age group of the participants. The within-subject variable was task type.

### *Procedure*

First, using structured questionnaires, the researchers recorded the participants' background information, including age, sex, height, weight, fall history in the past 6 months, education, smartphone experience, and internet experience. Fall history was determined by asking the question "Have you ever fallen unintentionally in the past six months?" Smartphone experience was determined by asking the question "Are you using a smartphone?" Internet experience was determined by asking the question "Do you use the internet?" For each participant, the Activity-Specific Balance Confidence Scale (ABC) [26] was used for measuring the fear of falling, the TUG test [27] was performed for measuring mobility, and the one leg stance (OLS) test [28] was performed for measuring balance.

Second, the participants were asked to take gait tests under the following three task types using the app: ST, DT, and FW conditions. The participants were initially asked to walk at a comfortable speed. Then, the participants were asked to walk at a comfortable speed while serially subtracting 3 from a 3-digit number randomly given by the experimenter, stating the answers out loud. Finally, the participants were asked to walk as fast as possible. All the participants wore comfortable footwear. The smartphone was placed in a waist-worn pocket, located close to the L3 region.

Third, participants completed the acceptability and usability survey with the researchers' help. The survey comprised 3 sections. The first section included 10 items measured on a 5-point Likert scale, with responses ranging from strongly agree to strongly disagree. The questionnaire used to measure acceptance was adopted from a study by Zhou et al [29]. There were four quantitative measures of acceptance (Multimedia Appendix 2): perceived ease of use, perceived usefulness, ease of learning, and intention to use. The second section included usability testing with the System Usability Scale (SUS). The SUS measures the usability of a product and consists of 10 items which are evaluated on a 5-point Likert scale ranging from 1 for *strongly disagree* to 5 for *strongly agree*. The results were distributed on a specific scale ranging from 0 for *worst imaginable* to 100 for *best imaginable* [30].

Finally, the participants were asked the following questions during the interviews: (1) What features of this app do you like? (2) What features of this app do you dislike? and (3) What do you think are the potential improvements for this app?

### *Statistical Analysis*

All statistical analyses were performed using IBM SPSS for Windows (version 22.0). Regarding the demographics and functional performance of the participants, normality was assessed using the Kolmogorov-Smirnov test. Independent *t* tests were used for the measures that were distributed normally. The Mann-Whitney *U* tests were used for the measures that were not distributed normally. Pearson chi-square tests were used to analyze the difference in categorical variables (sex, fall history, education, smartphone owner, and internet user) between participants from Beijing and Chongqing.

The gait parameters collected by the smartphone were analyzed using repeated measures analysis of variance (ANOVA). The gait data were assessed using the Mauchly test of sphericity. If sphericity was violated, the Greenhouse-Geisser correction was made. If the task type or the interaction effects were significant, post hoc tests were performed using the least significant difference (LSD). The level of significance was set at $P<.05$.

The user evaluation of acceptability and usability of the app was analyzed using descriptive statistics and ANOVA. We examined the differences in responses by participant characteristics (age and region). The interview data about participants' recommendations were analyzed using content analysis. The open-ended responses were analyzed using magnitude coding, a process that quantifies participants' answers, highlighting the most frequent comments.

## Results

### Participants' Characteristics

The participants' characteristics are presented in Table 1. Most of the participants were female (108/148, 73.0%). The mean age of the participants was 69.8 (SD 7.0) years, ranging from 60 to 87 years. We compared demographics and functional performance of participants from Beijing and Chongqing. The participants from Beijing had greater values for height, weight, education, being a smartphone owner, and for being an internet user than participants from Chongqing ($P<.05$). Participants from Beijing also performed better in ABC, TUG, and OLS ($P<.001$).

**Table 1.** Demographics and functional performance of participants in this study (N=148).

| Variables | All (N=148) | Beijing (n=70) | Chongqing (n=78) | *P* value |
|---|---|---|---|---|
| Age (years), mean (SD) | 69.8 (7.0) | 70.5 (7.7) | 69.2 (6.2) | .27 |
| **Sex, n** | | | | **.06** |
| Male | 40 | 24 | 16 | |
| Female | 108 | 46 | 62 | |
| Height (cm), mean (SD) | 159.4 (8.2) | 162.2 (7.4) | 156.7 (8.1) | <.001 |
| Weight (kg), mean (SD) | 62.5 (12.8) | 65.1 (15.5) | 60.5 (9.2) | .02 |
| **Fall history in the past 6 months[a], n** | | | | **.85** |
| Yes | 37 | 18 | 19 | |
| No | 111 | 52 | 59 | |
| **Education, n** | | | | **<.001** |
| Primary | 38 | 1 | 37 | |
| Middle school | 54 | 20 | 34 | |
| High school or technical secondary school | 28 | 22 | 6 | |
| College or junior college | 28 | 27 | 1 | |
| **Smartphone owner[b], n** | | | | **<.001** |
| Yes | 109 | 64 | 45 | |
| No | 39 | 6 | 33 | |
| **Internet user[c], n** | | | | **<.001** |
| Yes | 89 | 54 | 35 | |
| No | 59 | 16 | 43 | |
| Activity-Specific Balance Confidence scale (%), mean (SD) | 89.7 (10.2) | 92.4 (8.8) | 87.0 (10.8) | .001 |
| Timed up and go (seconds), mean (SD) | 9.5 (2.1) | 8.5 (1.4) | 10.3 (2.2) | <.001 |
| One leg stance (seconds), mean (SD) | 21.2 (9.8) | 25.7 (7.2) | 17.9 (10.3) | <.001 |

[a]Fall history was determined by asking the question “Have you ever fallen unintentionally in the past six months?”

[b]Smartphone experience was determined by asking the question “Are you using a smartphone?”

[c]Internet experience was determined by asking the question “Do you use the internet?”

## Gait Assessment

After checking the data collected by the smartphone, the data of 8 participants were excluded because of abnormal collection (missing data). Therefore, the number of participants included for gait analysis was 140. Of the 8 excluded participants, 6 were from Beijing, and 2 were from Chongqing; 5 were in the age group 60 to 69 years, 1 was in the age group 70 to 79 years, and 2 were in the age group 80 to 89 years. In all, 3 of the excluded participants were male, and 5 were female.

### *Step Frequency*

Table 2 presents statistics for step frequency. ANOVA indicated that age group ($F_{2,134}$=0.45, $P$=.50) and task type ($F_{1.72,230.43}$=204.16, $P$<.001) had significant effects on step frequency. The step frequency was 2.07, 2.01, and 1.95 Hz for the age groups 60 to 69 years, 70 to 79 years, and 80 to 89 years, respectively. The post hoc analysis showed the step frequency of participants in the age group 60 to 69 years was significantly higher than that of participants in the other age groups.

Step frequency was 1.85, 2.18, and 1.99 Hz for task types DT, FW, and ST, respectively. The post hoc analysis showed that step frequencies in DT, FW, and ST were significantly different from each other ($P$ values<.001). The lowest step frequency was observed in DT, whereas the highest step frequency was observed in ST.

**Table 2.** Statistics for step frequency.

| Variables | Descriptive analysis, mean (95% CI) | Analysis of variance | |
|---|---|---|---|
| | | *F* (*df*) | *P* value |
| **Region** | | **0.45 (1)** | **.50** |
| Beijing (n=64) | 2.00 (1.95-2.04) | | |
| Chongqing (n=76) | 2.02 (1.97-2.07) | | |
| **Age group (years)** | | **4.61 (2)** | **.01**[a] |
| 60-69 (n=79) | 2.07 (2.03-2.10) | | |
| 70-79 (n=43) | 2.01 (1.96-2.05) | | |
| 80-89 (n=18) | 1.95 (1.88-2.03) | | |
| **Task type** | | **204.16 (1.72)** | **<.001**[a] |
| Dual task | 1.85 (1.81-1.89) | | |
| Fast walking | 2.18 (2.15-2.22) | | |
| Single task | 1.99 (1.96-2.02) | | |

[a]Significant at .05 level.

### *Acceleration Root Mean Square*

Table 3 presents statistics for RMS. Task type ($F_{1.78,237.84}$=302.94, $P$<.001) and region×task type ($F_{1.78,237.84}$=6.15, $P$=.004) were demonstrated to have significant effects on RMS.

For participants from Beijing, RMS in DT, FW, and ST was 1.97, 3.14, and 2.26 m/s$^2$, respectively. The post hoc analysis showed that RMS in DT, FW, and ST were significantly different from each other ($P$ values <.001). RMS was highest in FW and lowest in DT.

For participants from Chongqing, RMS in DT, FW, and ST was 2.05, 3.06, and 2.36 m/s$^2$, respectively. The post hoc analysis showed RMS in DT, FW, and ST were significantly different from each other ($P$ values <.001). RMS was highest in FW and lowest in DT.

**Table 3.** Statistics for root mean square.

| Variables | Descriptive analysis, mean (95% CI) | Analysis of variance | |
|---|---|---|---|
| | | *F* (*df*) | *P* value |
| **Region** | | **0.11 (1)** | **.75** |
| Beijing | 2.46 (2.31-2.60) | | |
| Chongqing | 2.49 (2.33-2.65) | | |
| **Age group (years)** | | **2.10 (2)** | **.13** |
| 60-69 | 2.53 (2.41-2.64) | | |
| 70-79 | 2.60 (2.44-2.76) | | |
| 80-89 | 2.29 (2.04-2.55) | | |
| **Task type** | | **302.94 (1.78)** | **<.001**[a] |
| DT[b] | 1.97 (1.85-2.09) | | |
| FW[c] | 3.14 (3.01-3.27) | | |
| ST[d] | 2.31 (2.20-2.42) | | |
| **Region×task type** | | **6.15 (1.78)** | **.004**[a] |
| DT (Beijing) | 1.89 (1.73-2.05) | | |
| FW (Beijing) | 3.22 (3.05-3.40) | | |
| ST (Beijing) | 2.26 (2.11-2.41) | | |
| DT (Chongqing) | 2.05 (1.87-2.23) | | |
| FW (Chongqing) | 3.06 (2.87-3.26) | | |
| ST (Chongqing) | 2.36 (2.20-2.52) | | |

[a]Significant at .05 level.

[b]DT: dual task.

[c]FW: fast walking.

[d]ST: single task.

### *Step Variability*

Table 4 presents statistics for step variability. Task type ($F_{1.59,212.89}$=16.77, $P$<.001) and age×task type ($F_{3.18,212.89}$=3.57, $P$=.01) were demonstrated to have significant effects on RMS.

For participants in the age group 60 to 69 years, step variability in DT, FW, and ST was 0.093, 0.13, and 0.093, respectively. Step variability in DT was significantly lower than that in FW ($P$<.001). Step time variability in FW was significantly higher than that in ST ($P$<.001).

For participants in the age group 70 to 79 years, step variability in DT, FW, and ST was 0.105, 0.121, and 0.092, respectively. Step variability in FW was significantly higher than that in ST ($P$<.001).

For participants in the age group 80 to 89 years, step variability in DT, FW, and ST was 0.098, 0.091, and 0.077, respectively. Step variability in DT was significantly higher than that in ST ($P$=.003).

**Table 4.** Statistics for step variability.

| Variables | Descriptive analysis, mean (95% CI) | Analysis of variance | |
|---|---|---|---|
| | | *F* (*df*) | *P* value |
| **Region** | | **3.83 (1)** | **.05** |
| Beijing | 0.091 (0.080-0.10) | | |
| Chongqing | 0.11 (0.095-0.12) | | |
| **Age group (years)** | | **1.20 (2)** | **.30** |
| 60-69 | 0.10 (0.095-0.11) | | |
| 70-79 | 0.11 (0.094-0.12) | | |
| 80-89 | 0.089 (0.069-0.11) | | |
| **Task type** | | **16.77 (1.59)** | **<.001[a]** |
| DT[b] | 0.099 (0.090-0.11) | | |
| FW[c] | 0.11 (0.10-0.12) | | |
| ST[d] | 0.087 (0.079-0.09)6 | | |
| **Age group×task type** | | **3.57 (3.18)** | **.01[a]** |
| 60-69 (DT) | 0.093 (0.083-0.10) | | |
| 70-79 (DT) | 0.11 (0.091-0.12) | | |
| 80-89 (DT) | 0.098 (0.077-0.12) | | |
| 60-69 (FW) | 0.13 (0.11-0.14) | | |
| 70-79 (FW) | 0.12 (0.11-0.14) | | |
| 80-89 (FW) | 0.091 (0.065-0.12) | | |
| 60-69 (ST) | 0.093 (0.084-0.10) | | |
| 70-79 (ST) | 0.092 (0.079-0.11) | | |
| 80-89 (ST) | 0.077 (0.056-0.097) | | |

[a]Significant at .05 level.

[b]DT: dual task.

[c]FW: fast walking.

[d]ST: single task.

### *Step Regularity*

Table 5 presents statistics for step regularity. Region ($F_{1,134}$=4.51, *P*=.04), task type ($F_{1.45,194.66}$=57.30, *P*<.001), and age group×task type ($F_{2.91,194.66}$=7.02, *P*<.001) had significant effects on step regularity. Step regularity for participants from Beijing and Chongqing was 0.75 and 0.79, respectively. Participants from Beijing had significantly lower step regularity than participants from Chongqing (*P*=.04).

For participants in the age group 60 to 69 years, step regularity in DT, FW, and ST was 0.76, 0.79, and 0.81, respectively. The post hoc analysis showed that step regularity values in DT, FW, and ST were significantly different from each other (*P* values <.05). Step regularity was highest in ST and lowest in DT.

For participants in the age group 70 to 79 years, step regularity in DT, FW, and ST was 0.71, 0.80, and 0.80, respectively. Step regularity in DT was significantly lower than that in FW (*P*<.001) and in ST (*P*<.001).

For participants in the age group 80 to 89 years, step regularity in DT, FW, and ST was 0.67, 0.83, and 0.79, respectively. Step regularity values in DT, FW, and ST were significantly different from each other (*P* values <.05).

**Table 5.** Statistics for step regularity.

| Variables | Descriptive analysis, mean (95% CI) | Analysis of variance | |
|---|---|---|---|
| | | *F* (*df*) | *P* value |
| **Region** | | **4.51 (1)** | **.04**[a] |
| Beijing | 0.75 (0.73-0.78) | | |
| Chongqing | 0.79 (0.77-0.82) | | |
| **Age group (years)** | | **0.69 (2)** | **.50** |
| 60-69 | 0.79 (0.77-0.81) | | |
| 70-79 | 0.77 (0.74-0.80) | | |
| 80-89 | 0.76 (0.72-0.81) | | |
| **Task type** | | **57.30 (1.45)** | **<.001**[a] |
| DT[b] | 0.72 (0.69-0.74) | | |
| FW[c] | 0.81 (0.79-0.83) | | |
| ST[d] | 0.80 (0.78-0.82) | | |
| **Age group×task type** | | **7.02 (2.91)** | **<.001**[a] |
| 60-69 (DT) | 0.76 (0.73-0.79) | | |
| 70-79 (DT) | 0.71 (0.68-0.75) | | |
| 80-89 (DT) | 0.67 (0.61-0.73) | | |
| 60-69 (FW) | 0.79 (0.77-0.82) | | |
| 70-79 (FW) | 0.80 (0.77-0.83) | | |
| 80-89 (FW) | 0.83 (0.78-0.88) | | |
| 60-69 (ST) | 0.81 (0.78-0.83) | | |
| 70-79 (ST) | 0.80 (0.77-0.83) | | |
| 80-89 (ST) | 0.79 (0.74-0.84) | | |

[a]Significant at .05 level.

[b]DT: dual task.

[c]FW: fast walking.

[d]ST: single task.

### *Step Symmetry*

Table 6 presents the statistics for step symmetry. Task type ($F_{1.60,214.51}$=13.52, $P$<.001) had significant effects on step symmetry. Step symmetry in DT, FW, and ST was 0.89, 0.93, and 0.93, respectively. Step symmetry in DT was significantly lower than that in FW ($P$<.001) and in ST ($P$<.001).

**Table 6.** Statistics for step symmetry.

| Variables | Descriptive analysis, mean (95% CI) | Analysis of variance | |
|---|---|---|---|
| | | *F* (*df*) | *P* value |
| **Region** | | **2.09 (1)** | **.15** |
| Beijing | 0.91 (0.88-0.93) | | |
| Chongqing | 0.93 (0.91-0.96) | | |
| **Age group (years)** | | **0.67 (2)** | **.51** |
| 60-69 | 0.92 (0.90-0.94) | | |
| 70-79 | 0.91 (0.88-0.93) | | |
| 80-89 | 0.93 (0.89-0.97) | | |
| **Task type** | | **13.52 (1.60)** | **<.001**[a] |
| Dual task | 0.89 (0.87-0.92) | | |
| Fast walking | 0.93 (0.91-0.95) | | |
| Single task | 0.93 (0.91-0.95) | | |

[a]Significant at .05 level.

## Acceptability

Overall, the acceptability feedback from users was positive for the four quantitative measures of acceptance (Table 7), indicating that participants acknowledged the perceived ease of use, perceived usefulness, ease of learning, and intention to use.

**Table 7.** Descriptive statistics for the acceptability of the app.

| Function | Perceived ease of use, mean (SD) | Perceived usefulness, mean (SD) | Ease of learning, mean (SD) | Intention to use, mean (SD) |
|---|---|---|---|---|
| Gait test | 3.7 (0.6) | 3.9 (0.4) | 3.7 (0.6) | 3.5 (0.8) |
| Viewing graph | 3.6 (0.8) | 3.8 (0.5) | 3.6 (0.7) | 3.6 (0.7) |
| Viewing report | 3.5 (0.7) | 3.8 (0.5) | 3.6 (0.7) | 3.5 (0.8) |

## Usability

The data of 2 participants were excluded because of missing data. Therefore, the number of participants included for usability analysis was 146. Participants identified the usability of the system with an overall SUS score of 59.7 (SD 10.7) out of 100. In terms of sex, there was no noticeable difference for the perception of usability. Male participants evaluated the usability of the system with a score of 62.1 (SD 11.6), whereas female participants evaluated the usability with a score of 58.8 (SD 10.3).

Regarding region, there was a significant difference in the SUS score between participants in Beijing and Chongqing ($t_{144}$=4.17, *P*<.001), indicating that participants in Beijing (mean 63.4, SD 9.8) had a higher level of satisfaction with the gait assessment app compared with participants in Chongqing (mean 56.3, SD 10.5). A possible reason for this difference is that participants in Beijing have higher education than participants in Chongqing.

Moreover, there was a significant difference in the SUS score between age groups ($F_{2,145}$=3.09, *P*=.048). The LSD post hoc revealed that participants between 60 and 69 years of age had a higher level of satisfaction (mean 61.4, SD 10.1) than participants over 80 years of age (mean 55.3, SD 11.7; *P*=.02).

## Attitude Toward Key Features of the App

### *Question 1: What Features of This App Do You Like?*

During the interviews, participants reported the app features they liked. Of the 140 participants, 48 (34.3%) thought that the app could help them familiarize themselves with their health conditions, and older adults could benefit from the app both physically and mentally:

> *You can know the condition of how you walk in your daily life and see the results at a glance. You can plan how to walk in your own life and walking would become a daily fun.*
>
> *It is good to see the speed and balance during walking. If the result is good, the mood is good.*
>
> *It is good for health, helping the elderly to train their brain.*

In all, 21 of 140 (15.0%) participants considered the app to be convenient to use:

> *I think the application is convenient.*
>
> *It is easy to use the application. I will know about the situation at a glance.*
>
> *It is flexible. You can use it anytime and anywhere as long as you have a mobile phone.*

Of the 140 participants, 6 (4.3%) thought that the app gave objective indicators that could not be observed by eye:

> *I like the authenticity of the app. Slow is slow.*
>
> *The app reflects the physical condition objectively.*

Of the 140 participants, 5 (3.5%) expressed the intention to use the app as an incentive for walking or exercise:

> *Walking is good. I like walking.*
>
> *It urges me to take exercise.*
>
> *The application reflected the whole process of walking, which could help you exercise in a way that suits you.*
>
> *The application could offer scientific quantitative analysis, which is good for walking.*

### *Question 2: What Features of This App Do You Dislike?*

Next, participants also reported the app features that they disliked, expressing difficulties of using the app. Of 140 participants, 37 (26.4%) said that there was nothing they did not like about the app. Of 140 participants, 25 (17.9%) complained about the small font size on the display, as one of them mentioned:

> *The font size of the application is too small to see clearly.*

Of 140 participants, 14 (10.0%) older adults thought that using the app was too complicated for them to learn and mentioned that "The system is a little bit too complicated," but 3 of them showed that they could learn to use the app with the help of others (eg, a teacher or daughter).

In all, 5.7% (8/140) participants complained about the graph, as one of them pointed out:

> *I don't understand this graph. It is better to explain which range is right/good and which range is wrong/bad. It is better to have a normal range.*

Of 140 participants, 5 (3.6%) thought that the gait indicators in the report were difficult to understand. Some participants thought that the result should be saved and recalled:

> *It's very difficult to learn. I remembered it at that time and then I would forget it. I walk if I have to, and I don't have to look at it. I use mobile phone for the elderly, so I do not know about this.*

In all, 3.5% (5/140) participants thought that the functions of the app should be supplemented. One participant mentioned:

> *It would be better to include blood pressure and heart rate. When exercising, elderly people are required to reach a state of slight sweating. The app could be a reminder of excessive exercise and be used to detect the warning signs of sudden death during sports.*

Another participant said:

> *The function is relatively simple if it simply measures walking. Walking may be good, but the balance system may not be good. It is better to give some guidelines on how to improve the balance system.*

Of 140 participants, 3 (2.1%) regarded the device restrictions as disadvantages:

> *The application has equipment restrictions. It will be better if it could be used in a mobile phone for the elderly.*

Overall, 2 of 140 (1.4%) participants thought it was unnecessary to use the app by giving the following reasons:

> *I rarely go out.*
>
> *There is no big change in the pace when walking on the ground. If it is serious, the doctor can diagnose it. If it is not serious, the app will not reflect.*

One participant complained that the pocket was too large, while another participant said the following:

> *The size of the smartphone was too large. It would be better to be as small as a smart bracelet.*

These responses from the participants explained the usability problems and reflected future directions for improvement.

### *Question 3: What Are the Potential Improvements for This App?*

Finally, participants identified various potential improvements for the app. Of 140 participants, 35 (25.0%) suggested that the font size of the app should be larger, and 37 (26.4%) suggested that the functions of the app should be supplemented. Among them, 11 participants suggested that the app should provide more information on the purpose of gait test:

> *What is the purpose of each test? Does it inform what problem the body has? What does the value of each gait parameter inform? How to improve if the value is higher or lower?*

In all, 13 of 140 (9.3%) participants suggested that the app should provide reference values for gait parameters so that the user knows which position they are in:

> *Adding the reference range of gait parameters. Tell users which range is preferred.*
>
> *The results must be accurate. Not only must there be results, but there must also be explanations for the reasons of the results, and some medical help.*

Regarding the graph, of 140 participants, 1 (0.7%) indicated:

> *It would be better for the chart to display a curve and for long-term users to be able to provide the data to the doctor to assess the condition, which would allow to doctor to make informed judgements.*

Overall, 2 of 140 (1.4%) participants suggested using distinct colors:

> *It is better to use differentiated colors. It is clearer at a glance.*

Some participants suggested that the app should include other body indicators during walking, such as heart rate (11 participants), blood pressure (5), step count (4), vital capacity (1), and distance (1). Some suggested that the app could add other functions, such as the option for background music (2), road condition prompts (2), other exercise activities (1), bone density test (1), and viewing the results of friends (1).

Of 140 participants, 11 (7.9%) thought that the functions of the app should be simplified:

> *I hope there is a summary of the result and the app informs me whether the result is good or bad. There could be fewer gait indicators.*

In all, 4 of 140 (2.9%) participants preferred other wearing positions than the lower back:

*I hope the smart phone could be placed on any part of the body, such as on the chest. Placing the smartphone in the pocket of the suit is more convenient than on the back.*

Of 140 participants, 2 (1.4%) suggested changing the environment when conducting the test:

> *The test could be conducted in an open environment. It is not as enjoyable to walk in a closed environment as it is outdoors.*

One (1/140, 0.7%) participant hoped that the app could be used in a mobile phone for the elderly.

# Discussion

## Principal Findings

We developed a gait-monitoring mobile phone app (Pocket Gait) and assessed the acceptability of the app with older adults. We examined the influence of age group, task type, and region on gait parameters. Gait assessment results revealed that task type had a significant effect on all gait parameters, namely, step frequency, RMS, step variability, step regularity, and step symmetry (all $P$ values <.001). Age had a significant effect on step frequency ($P$=.01), and region had a significant effect on step regularity ($P$=.04). The step frequency of participants in the age group 60 to 69 years was significantly higher than that of participants in the other age groups. Participants from Beijing had significantly lower step regularity than participants from Chongqing.

We performed acceptability and usability testing among participants. The acceptability of the gait-monitoring app was positive among older adults for the four quantitative measures of acceptance, namely, the perceived ease of use, perceived usefulness, ease of learning, and intention to use. The usability score was 59.7 out of 100. Further interviews indicated some usability problems. Suggestions to improve the usability of the app are presented in Textbox 1. Basic improvements suggested are that the font size should be larger, and more detailed instructions should be provided. To reduce difficulties of using the app, users should be provided with instructions and training, and informed about the meaning of gait parameters and the use of each test when promoting the mobile phone app.

**Textbox 1.** Suggestions for improvement.

*Gait test*

1. The app should inform users about the purpose of each gait test.
2. Inclusion of heart rate and blood pressure monitoring in the app.
3. Use higher volume for the voice instructions.
4. The gait test could be conducted in an open environment.
5. It is more comfortable to walk if background music is added.
6. The smartphone could be placed on any part of the body, such as on the chest.
7. Tell the user what range of the gait parameters is acceptable.
8. Change *test* to *activity* (or other words) to relax the users.

*Viewing graph*

9. For the long term, users can provide the data to a doctor to assess the condition.
10. Save the chart and recall it at any time.

*Viewing report*

11. Use larger font size.
12. Add the reference range for gait parameters and tell users which range is preferred.
13. Present results, as well as explanation for the results, and some medical advice.
14. Tell the user how to improve if the value is higher or lower than the reference value.

## Comparison With Previous Work

Other studies in health-related ICT also applied acceptability and usability testing among older adults. Portz et al [31] developed a mobile phone app for tracking symptoms of heart failure among older adults and tested its acceptability. The study found that older age was associated with a need for assistance to use the app. Vaziri et al's study [32,33] also found that age was an important factor for the system usability evaluation of an ICT-based fall prevention system iStoppFalls. Younger participants assessed the usability of the system better than older participants [33]. Our study showed that age and region are important factors for the usability assessment of the gait assessment app. Participants that were younger and from Beijing assessed the usability of the app better. During the experiment, we found that there were some difficulties for the older adults

aged over 80 years, for example, eye disease prevented them from reading the words on the screen. In addition, 39/148 (26.4%) of the older adults in this study did not have a smartphone, so that they had no idea about the app. Therefore, it will be important to take factors such as age and region into consideration when promoting the mobile phone-based gait assessment app. Older users or users with low socioeconomic status may be disadvantaged in using the app. More instructions and social support from the caregivers or family members are needed to promote using the app among such users.

### Limitations and Future Studies

The study has some limitations. First, the study only included community-dwelling older adults from Beijing and Chongqing. Older adults from other regions were not investigated in this study. Second, the gait assessment was conducted in a corridor, which was different from their daily environment. Third, the researchers were helping the participants complete the usability and acceptability testing. If the participant could not read, the researcher would read aloud each item of acceptability and usability. It is worth noting that Chinese older adults tend to give moderate responses during the user evaluation, that is, they seldom responded *strongly disagree* or *strongly agree*. This could partly explain why the SUS score is marginal.

The gait assessment app could be generalized to other populations. For example, it could be an incentive for exercise for sedentary young people. The app could also be used to monitor long-term changes in patients undergoing rehabilitation (eg, stroke and Parkinson disease). As this study evaluated the gait assessment app from the perspective of community-dwelling older adults, future studies could evaluate the app from the perspective of patients or care givers. Finally, the app could be extended to other mobile platforms (eg, iPhone Operating System).

### Conclusions

Smartphones may serve as useful tools to support the gait assessment of older adults and facilitate aging in place, which is defined as “remaining living in the community, with some level of independence, rather than in residential care” [34]. Our study discussed the development and acceptability of a gait-monitoring mobile phone app (Pocket Gait) among Chinese older adults. The study findings established reference values for gait parameters and provided design recommendations for further improvements of the app.

## Acknowledgments

This study was funded by the National Natural Science Foundation of China (grant number 71661167006) and the Fundamental Research Funds for the Central Universities (grant number JUSRP12068). The authors thank Ye Yuan for his help in Android programming, Dr Wei Zhao for his help in conducting the experiment, and Prof Jia Zhou for her help in recruiting participants and conducting the experiment.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
The algorithm to extract gait parameters.
[DOCX File , 13 KB - mhealth_v8i5e14453_app1.docx ]

Multimedia Appendix 2
Questionnaires for acceptability and usability.
[DOCX File , 19 KB - mhealth_v8i5e14453_app2.docx ]

## References

1. National National Bureau of Statistics of China. The population has grown steadily, and the population quality has improved significantly. 2019 Aug 22. URL: http://www.stats.gov.cn/tjsj/zxfb/201908/t20190822_1692898.html [accessed 2020-01-12]
2. World Health Organization. WHO Global Report on Falls Prevention in Older Age. Geneva: World Health Organization; 2007.
3. Paulson S, Gray M. Parameters of gait among community-dwelling older adults. J Geriatr Phys Ther 2015;38(1):28-32. [doi: 10.1519/JPT.0000000000000018] [Medline: 24755689]
4. Hausdorff JM, Rios DA, Edelberg HK. Gait variability and fall risk in community-living older adults: a 1-year prospective study. Arch Phys Med Rehabil 2001 Aug;82(8):1050-1056. [doi: 10.1053/apmr.2001.24893] [Medline: 11494184]
5. Moe-Nilssen R, Helbostad JL. Interstride trunk acceleration variability but not step width variability can differentiate between fit and frail older adults. Gait Posture 2005 Feb;21(2):164-170. [doi: 10.1016/j.gaitpost.2004.01.013] [Medline: 15639395]
6. Montero-Odasso M, Muir SW, Hall M, Doherty TJ, Kloseck M, Beauchet O, et al. Gait variability is associated with frailty in community-dwelling older adults. J Gerontol A Biol Sci Med Sci 2011 May;66(5):568-576. [doi: 10.1093/gerona/glr007] [Medline: 21357190]

7. Beauchet O, Blumen HM, Callisaya ML, de Cock A, Kressig RW, Srikanth V, et al. Spatiotemporal gait characteristics associated with cognitive impairment: a multicenter cross-sectional study, the intercontinental 'Gait, cognition & decline' initiative. Curr Alzheimer Res 2018 Jan 23;15(3):273-282. [doi: 10.2174/1567205014666170725125621] [Medline: 28745226]
8. Menz HB, Lord SR, Fitzpatrick RC. Age-related differences in walking stability. Age Ageing 2003 Mar;32(2):137-142. [doi: 10.1093/ageing/32.2.137] [Medline: 12615555]
9. Kosse NM, Vuillerme N, Hortobágyi T, Lamoth CJ. Multiple gait parameters derived from iPod accelerometry predict age-related gait changes. Gait Posture 2016 May;46:112-117. [doi: 10.1016/j.gaitpost.2016.02.022] [Medline: 27131187]
10. Montero-Odasso M, Casas A, Hansen KT, Bilski P, Gutmanis I, Wells JL, et al. Quantitative gait analysis under dual-task in older people with mild cognitive impairment: a reliability study. J Neuroeng Rehabil 2009 Sep 21;6:35 [FREE Full text] [doi: 10.1186/1743-0003-6-35] [Medline: 19772593]
11. Smith E, Walsh L, Doyle J, Greene B, Blake C. Effect of a dual task on quantitative Timed Up and Go performance in community-dwelling older adults: A preliminary study. Geriatr Gerontol Int 2017 Aug;17(8):1176-1182. [doi: 10.1111/ggi.12845] [Medline: 27427514]
12. Schwenk M, Howe C, Saleh A, Mohler J, Grewal G, Armstrong D, et al. Frailty and technology: a systematic review of gait analysis in those with frailty. Gerontology 2014;60(1):79-89 [FREE Full text] [doi: 10.1159/000354211] [Medline: 23949441]
13. Zhong R, Rau PP, Yan X. Application of smart bracelet to monitor frailty-related gait parameters of older Chinese adults: a preliminary study. Geriatr Gerontol Int 2018 Sep;18(9):1366-1371. [doi: 10.1111/ggi.13492] [Medline: 30105810]
14. Blumenthal J. A Mobile Application for Dual-Task Gait Assessment of Cognitive Status. Canada: University of Toronto; 2017. URL: https://pdfs.semanticscholar.org/cda3/43decf71f4bf34c463bf407080c1bff6821a.pdf [accessed 2020-01-15]
15. Manor B, Yu W, Zhu H, Harrison R, Lo O, Lipsitz L, et al. Smartphone app-based assessment of gait during normal and dual-task walking: demonstration of validity and reliability. JMIR Mhealth Uhealth 2018 Jan 30;6(1):e36 [FREE Full text] [doi: 10.2196/mhealth.8815] [Medline: 29382625]
16. Miura T, Yabu K, Hiyama A, Inamura N, Hirose M, Ifukube T. Smartphone-Based Gait Measurement Application for Exercise and Its Effects on the Lifestyle of Senior Citizens. In: Proceedings of the IFIP Conference on Human-Computer Interaction. 2015 Presented at: INTERACT'15; September 14-18, 2015; Bamberg, Germany. [doi: 10.1007/978-3-319-22698-9_7]
17. Wagner R, Ganz A. PAGAS: Portable and Accurate Gait Analysis System. In: Proceedings of the 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2012 Presented at: EMBC'12; August 28 - September 1, 2012; San Diego, CA, USA p. 280-283. [doi: 10.1109/embc.2012.6345924]
18. Nishiguchi S, Yamada M, Nagai K, Mori S, Kajiwara Y, Sonoda T, et al. Reliability and validity of gait analysis by android-based smartphone. Telemed J E Health 2012 May;18(4):292-296. [doi: 10.1089/tmj.2011.0132] [Medline: 22400972]
19. Liu J, Ying L, Rau P, Hui L, Xia W, Li D. How Socio-Economic Structure Influences Rural Users' Acceptance of Mobile Entertainment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010 Presented at: CHI'10; April 10 - 15, 2010; Atlanta, Georgia, USA, p. 2203-2212. [doi: 10.1145/1753326.1753659]
20. Shen C, Wang MP, Chu JT, Wan A, Viswanath K, Chan SS, et al. Health app possession among smartphone or tablet owners in Hong Kong: population-based survey. JMIR Mhealth Uhealth 2017 Jun 5;5(6):e77 [FREE Full text] [doi: 10.2196/mhealth.7628] [Medline: 28583905]
21. National Bureau of Statistics of China. Internet. China Statistical Yearbook 2018 URL: http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm [accessed 2020-01-13]
22. Zhong R, Rau PL, Yan X. Gait assessment of younger and older adults with portable motion-sensing methods: a user study. Mob Inf Syst 2019;2019:1-13 [FREE Full text] [doi: 10.1155/2019/1093514]
23. healthline. L3 URL: https://www.healthline.com/human-body-maps/l3-third-lumbar-spine-vertebrae#1 [accessed 2020-01-15]
24. Statista. 2019. Market Share of Mobile Operating Systems in China From January 2013 to August 2019 URL: https://www.statista.com/statistics/262176/market-share-held-by-mobile-operating-systems-in-china/ [accessed 2020-01-15]
25. Android Developers. SensorManager URL: https://developer.android.com/reference/android/hardware/SensorManager.html#SENSOR_DELAY_FASTEST [accessed 2020-01-15]
26. Guan Q, Han H, Li Y, Zhao L, Jin L, Zhan Q. Activities-specific Balance Confidence (ABC) Scale adapted for the mainland population of China. Clin Rehabil 2012 Jul;26(7):648-655. [doi: 10.1177/0269215511427748] [Medline: 22169829]
27. Podsiadlo D, Richardson S. The timed 'Up & Go': a test of basic functional mobility for frail elderly persons. J Am Geriatr Soc 1991 Feb;39(2):142-148. [doi: 10.1111/j.1532-5415.1991.tb01616.x] [Medline: 1991946]
28. Vellas BJ, Wayne SJ, Romero L, Baumgartner RN, Rubenstein LZ, Garry PJ. One-leg balance is an important predictor of injurious falls in older persons. J Am Geriatr Soc 1997 Jun;45(6):735-738. [doi: 10.1111/j.1532-5415.1997.tb01479.x] [Medline: 9180669]
29. Zhou J, Rau PP, Salvendy G. Older adults' text entry on smartphones and tablets: investigating effects of display size and input method on acceptance and performance. Int J Hum-Comput Interact 2014;30(9):727-739. [doi: 10.1080/10447318.2014.924348]

30. Brooke J. Hell - Jens Oliver Meiert. SUS-A Quick and Dirty Usability Scale URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2020-01-15]
31. Portz JD, Vehovec A, Dolansky MA, Levin JB, Bull S, Boxer R. The development and acceptability of a mobile application for tracking symptoms of heart failure among older adults. Telemed J E Health 2018 Feb;24(2):161-165 [FREE Full text] [doi: 10.1089/tmj.2017.0036] [Medline: 28696832]
32. Vaziri DD, Aal K, Gschwind YJ, Delbaere K, Weibert A, Annegarn J, et al. Analysis of effects and usage indicators for a ICT-based fall prevention system in community dwelling older adults. Int J Hum-Comput Stud 2017;106:10-25. [doi: 10.1016/j.ijhcs.2017.05.004]
33. Vaziri DD, Aal K, Ogonowski C, von Rekowski T, Kroll M, Marston HR, et al. Exploring user experience and technology acceptance for a fall prevention system: results from a randomized clinical trial and a living lab. Eur Rev Aging Phys Act 2016;13:6 [FREE Full text] [doi: 10.1186/s11556-016-0165-z] [Medline: 27293489]
34. Davey JA, de Joux V, Nana G, Arcus M. Accommodation Options for Older People in Aotearoa/New Zealand. Aotearoa New Zealand: Centre for Housing Research; 2004.

## Abbreviations

**ABC:** Activity-Specific Balance Confidence scale
**ANOVA:** analysis of variance
**CNY:** Chinese Yuan
**DT:** dual task
**FW:** fast walking
**GDP:** gross domestic product
**ICT:** information and communication technology
**L3:** the third lumbar spine vertebra
**LSD:** least significant difference
**OLS:** one leg stance
**RMS:** root mean square
**ST:** single task
**SUS:** System Usability Scale
**TUG:** timed up and go

*Edited by G Eysenbach; submitted 19.04.19; peer-reviewed by J Zhou, SF Liang, J Zhou, D Koutsouris, D Leightley, F Lanfranchi; comments to author 30.09.19; revised version received 19.11.19; accepted 16.12.19; published 26.05.20.*

*Please cite as:*
*Zhong R, Rau PLP*
*A Mobile Phone–Based Gait Assessment App for the Elderly: Development and Evaluation*
*JMIR Mhealth Uhealth 2020;8(5):e14453*
*URL: https://mhealth.jmir.org/2020/5/e14453*
*doi:10.2196/14453*
*PMID:32473005*

Original Paper

# A Lower Leg Physical Activity Intervention for Individuals With Chronic Venous Leg Ulcers: Randomized Controlled Trial

Teresa J Kelechi[1*], RN, PhD; Margaret A Prentice[1*], MBA; Martina Mueller[1,2*], PhD; Mohan Madisetti[1*], MSc; Alexey Vertegel[3*], PhD

[1]College of Nursing, Medical University of South Carolina, Charleston, SC, United States

[2]Department of Public Health Sciences, Medical University of South Carolina, Charleston, SC, United States

[3]Department of Bioengineering, Clemson University, Clemson, SC, United States

[*]all authors contributed equally

**Corresponding Author:**
Teresa J Kelechi, RN, PhD
College of Nursing
Medical University of South Carolina
99 Jonathan Lucas Street
MSC 160
Charleston, SC, 29425
United States
Phone: 1 843 792 4602
Email: kelechtj@musc.edu

## Abstract

**Background:** Individuals with venous leg ulcers (VLUs) suffer disproportionately with multiple chronic conditions, are often physically deconditioned, and demonstrate high levels of physical inactivity.

**Objective:** The primary objective of this randomized controlled trial was to establish the feasibility of a mobile health (mHealth) physical activity exercise app for individuals with VLUs to improve lower leg function.

**Methods:** In a 6-week study, adults with VLUs were recruited from 2 wound centers in South Carolina, United States, and enrolled if they were aged 18 years or older with impaired functional mobility and an ankle-brachial index between 0.8 and 1.3. Participants were randomized 1:1 to receive evidence-based, phased, nonexertive physical conditioning activities for lower leg function (FOOTFIT) or FOOTFIT+ with an added patient-provider communication feature. The mHealth Conditioning Activities for Lower Leg Function app also provided automated educational and motivational messages and user reports. Foot movement on the VLU-affected leg was tracked by a Bluetooth-enabled triaxial accelerometer. The study was guided by the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework to assess the feasibility of reach, adherence, acceptability, implementation, and maintenance.

**Results:** A total of 24 patients were recruited, enrolled, and randomized in the study. Most patients reported difficulty following the protocol for exercising and using the accelerometer and mobile phone and did not use the provider contact feature. However, all patients were adherent to the 6-week exercise program more than 85% of the time for duration, whereas 33% (8/24) of patients adhered more than 85% for the frequency of performing the exercises. Across the three exercise levels, adherence did not differ between the two groups. Confidence limits around the difference in proportions ranged from −0.4 to 0.7. Providers in FOOTFIT+ were inconsistent in checking participant progress reports because of lack of time from competing work commitments. The technology became outdated quickly, making maintenance problematic. Participants said they would continue to exercise their foot and legs and liked being able to follow along with the demonstrations of each level of exercise provided through the app.

**Conclusions:** The findings of this study suggest that despite initial interest in using the app, several components of the program as originally designed had limited acceptability and feasibility. Future refinements should include the use of more modern technology including smaller wearable accelerometers, mobile phones or tablets with larger screens, an app designed with larger graphics, automated reporting for providers, and more engaging user features.

**Trial Registration:** ClinicalTrials.gov NTC02632695; https://clinicaltrials.gov/ct2/show/NCT02632695

***(JMIR Mhealth Uhealth 2020;8(5):e15015)*** doi:10.2196/15015

**KEYWORDS**

leg ulcer; physical activity; exercise; mHealth; adherence; randomized controlled trial; feasibility

## Introduction

### Background

Individuals with venous leg ulcers (VLUs) suffer disproportionately with multiple chronic conditions and demonstrate high levels of inactivity [1,2]. Many patients are physically deconditioned and minimally ambulatory, able to only take a few steps at a time, are slow walkers, and have poor standing balance [3]. Obesity, older age, and leg pain are common characteristics that negatively impact functional abilities [4]. Reduced range of motion of the ankle and decreased calf muscle contractility with increased muscle deoxygenation are known physical impairments that also contribute to the worsening condition of the lower legs and substantially further restrict mobility [5]. These processes also contribute to poor wound healing outcomes.

Although numerous study findings suggest physical activity is important for improving outcomes in patients with ulcers, there are inconsistent findings about which types of programs are feasible for functionally impaired individuals with multiple chronic conditions. For those with leg wounds from venous or arterial diseases, enhanced healing and physical and functional abilities are key outcomes of physical activity programs. Evidence from a systematic review of six resistance exercise programs combined with compression therapy, the mainstay of treatment, failed to demonstrate improvements in the proportion of ulcers healed, quality of life, ankle range of motion, and calf muscle pump function [6]. The authors of the review reported the quality of evidence was low because of bias and imprecision of study methodology. Other problems with study methodology include patient and wound complexity, confounders such as comorbid conditions, the need for large samples sizes to show clinically relevant benefits, and the necessary long duration of the trials for wounds to heal, leading to high costs for high-quality trials [7]. However, data from studies of different types of physical activity such as supervised and self-management programs that include aerobic, resistance, and flexibility exercises specific for individuals with VLUs demonstrated high levels of feasibility, safety, and retention of participants [8,9]. Positive outcomes from these studies were noted in physical functions, including walking, sitting to standing, ankle range of movement, wound healing, pain, and lower ulcer recurrence rates. In a study in patients with VLUs, dorsiflexion exercises in which the calf muscle was pumped showed physiological improvements in skin perfusion, including oxygen content and blood flow [10]. Although limited research is available, the mechanisms by which exercises positively influence functional and healing outcomes are posited to be enhanced microvascular circulation through enhanced pumping function of the calf [10].

We previously conducted two small pilot studies of physical activity that informed the development of our study's intervention methods. The first was a Web-based physical activity intervention developed by our team of physical therapists and exercise specialists that included resistance bands, nonexertive foot movements, and a foot peddler (similar to peddling a bicycle), delivered by a coach (nursing student with degree in exercise science) through online face-to-face internet sessions [11]. A total of 5 participants with venous disease and a history of VLUs participated in determining the feasibility of engaging in three daily doses of nonexertive conditioning physical activities for lower leg function (CALF) for 1 week. We observed a very high level of patient satisfaction with working with the coach and using the equipment to engage in a variety of lower leg exercises while at home. We found the study procedures, including engagement using the Skype (Microsoft Corporation) interface, were feasible. Enhancements were made to CALF, one of which was the addition of a behavioral, motivational interviewing (MI) component, in our second 6-week study of 21 minimally ambulatory patients randomized to the CALF intervention or an exercise handout [12]. Certified wound care nurses were trained on MI communication techniques to interact with patients about engaging in exercises who were receiving wound care in a specialty clinic. We included only the nonexertive foot movements (did not use the peddler or resistance bands in this version of CALF) because of patients having ulcers and their lower legs being wrapped with multilayer compression that restricted ankle movement. The CALF intervention was found to be feasible and acceptable by both patients and the nurses who delivered it. However, having a coach and using providers who require special training are costly and not always available outside the clinic setting, as not all patients receive care in specialized wound clinics. Furthermore, many patients do not have access to physical therapy specialists or readily available transportation to attend exercise programs in the community, and because of having wounds and or functional impairments, many patients find it difficult to engage in physical activity.

### Objective

To address these barriers, technology-enhanced mobile health (mHealth) physical activity interventions are potential solutions to mitigate these barriers, as they have evolved for individuals with chronic conditions such as chronic obstructive pulmonary disease and provide availability for those with limited access to exercise programs [13,14]. In particular, accelerometer-based physical activity programs that provide feedback on steps taken, calories burned, and other physiologic data such as heart rate have become commonplace. Studies suggest that patients with chronic conditions want to participate in physical activity and would consider using an mHealth app if initiated through a clinic or medical office visit [15]. However, few studies integrate accelerometers with apps that routinely engage chronically ill patients and providers in a communication loop about progress toward goals and that are designed to promote adherence to physical activity [16,17].

To address this need, our team developed and tested a foot-based Bluetooth-enabled acceleration tracking (BEAT) device and mobile phone application system and demonstrated its reliability and validity in laboratory experiments using a standard

rotary-shaker test with 4 accelerometers (coefficient of variation was found to be 0.7%) and tested its feasibility in minimally ambulatory patients with venous disease and deconditioned legs [18]. The BEAT device detected even very minimal toe movements, which were captured by an app developed for mobile phones that received information from the accelerometer. The primary aim of this small randomized controlled trial study was to explore the feasibility of an exercise program comprised of an accelerometer-app combination, initiated during wound clinic visits, and performed by patients with VLUs in their homes.

## Methods

### Overview

This 6-week randomized clinical trial was designed to test the feasibility of a real-time, nonexertive, lower leg physical activity mHealth program, FOOTFIT, that combined BEAT and CALF. FOOTFIT was compared with FOOTFIT+ comprising BEAT, CALF, and connectivity to a wound clinician via text messaging, phone, or email. Although the primary goal was to establish feasibility, the intervention also targeted the function of the lower legs of individuals with VLUs. The trial complied with the Consolidated Standards of Research Trials (CONSORT) of Electronic and Mobile Health Applications and Online Telehealth guidelines [19], was approved by the Medical University of South Carolina institutional review board (#00043451), and was registered with ClinicalTrials.gov (NCT02632695) on December 17, 2015. Written informed consent was required to participate and obtained before enrollment in the study in which two visits occurred, one at baseline and one at 6 weeks. Participants received US $75 as compensation for participating in the study.

### Recruitment

Participants were recruited through direct referral from two participating wound clinics in the Southeastern region of the United States. Inclusion criteria were being aged 18 years and older, having a VLU, ankle-brachial index of 0.8 to 1.3 to rule out arterial insufficiency, receiving at least weekly wound care anticipated to last for at least 6 weeks from start of the study, being able to don the slipper onto which BEAT was affixed or having assistance from other, and being capable of using a mobile phone (individual observed using his or her phone after enrollment at baseline). Individuals were excluded if they had a comorbid condition such as stroke or severe arthritis that limited ankle function; an ulcer from other causes such as arterial, surgical, or traumatic; cognitive impairment determined by less than 3 recalled words and abnormal clock drawing on the MiniCog test [20] administered at baseline; or no 3G service available where the participant resided. Recruitment goals were established based on state population statistics according to US Census Bureau 2010 data: 3.86% (174,680/4,525,264) ethnic minorities, 28.52% (1,290,684/4,525,364) black, and 67.62% (3,060,000/4,525,364) % white.

### Sample Size and Randomization

The main outcome of this study was feasibility, and the sample size was determined based on pragmatic reasons. The aim was to recruit 24 participants who were randomized to FOOTFIT or FOOTFIT+ in a 1:1 ratio after baseline data collection. A computer-generated random number schema was developed by the statistician who had no contact with the participants. Group allocation was concealed in a database and revealed to the data collector after baseline data were collected and entered to minimize selection and measurement bias. Participants were then informed of allocation to either FOOTFIT or FOOTFIT+.

### FOOTFIT and FOOTFIT+ Interventions

The CALF program for this study consisted of 3 levels of phased, nonexertive movements for the lower legs, beginning with the most minimal level 1 intensity with foot on the floor progressing to level 2 intensity with heal on the floor and forefoot elevated, to maximal level 3 intensity with foot off the floor. Participants were instructed to perform each level for 2 weeks, 3 times per day for a minimum of 15 seconds each per activity, advancing in frequency and intensity. The BEAT was worn on the foot, affixed to a special slipper, during CALF to capture frequency and intensity of movements. The app on the mobile phone reminded individuals to perform the exercises at preset times each day, per patient preference, and sent patients supportive feedback after each daily session. There were also 12 short video clips of how to perform each movement and 16 short audio-recorded, evidenced-based information sessions about managing venous disease, the latest development in ulcer treatment, and other topics of interest expressed by individuals in our previous studies such as why compression is needed, what medications help healing, and how best to elevate the legs to reduce edema.

FOOTFIT+ was enhanced with an added phone, email, or text messaging connectivity feature to the wound care providers. The providers were instructed by the study principal investigator on theory-based patient-provider *talk* communication [21]. The providers were to make weekly contact with the participant via phone, email, or text, per participant preference, and provide a brief report of progress toward goals. If the prescribed activity was not being performed for 2 consecutive days by the participant, the providers were notified via text and reminded to contact the patient to verify the accelerometer was functioning and discuss any problems. The providers were informed that participants could contact them during normal business hours using any mode (phone, email, text, or voicemail), but the communication was to be related to the physical activity program only, such as pain during CALF and progress toward meeting goals. The quantity and quality of the interactions and whether the modes of communication were sustainable in actual practice were assessed. Participants in this group were also made aware that the providers would receive weekly reports about their progress.

### Measures and Outcomes

Data were collected preintervention at the baseline visit and at postintervention during the last visit week. Demographic information included health history, comorbid conditions, ulcer history, medications, age, sex, race/ethnicity, and rural/urban residence. Intervention feasibility assessment was guided by the RE-AIM framework for reach, effectiveness, adoption, adherence, acceptability, implementation, and maintenance

[22,23]. Reach was measured by continuous progress monitoring of sample representativeness, how patients learned about the study, types of recruitment activities and rates, meeting of recruitment goals (1 of 5 patients approached, and eligible would be consented), and number of eligible patients approached, consented, and oriented to the study. These data were captured on tracking forms, and quality checks were performed weekly and discussed at weekly team meetings. Data captured by the accelerometer BEAT was reported as the percentage of participants adherent to exercise duration defined as performing foot exercises at or above the recommended duration of 15 seconds or greater for each exercise (3 exercises per session, 3 times per day) or not always adherent (moved foot <15 seconds per exercise per session). Adherence to frequency was reported as the percentage of days the participants completed each CALF intensity level (3 levels performed 3 times per day, ie, 9 levels, each performed over 2 weeks); 9 levels performed 85% of the days or greater throughout the study period was considered adherent to frequency. Participants were instructed to perform the exercises 3 times daily, increase the duration of time each subsequent day, and review additional information on the app as needed. The number and reasons for dropouts were also recorded. Adoption focused on patient-provider communication and was measured by review of 10% of calls/emails/texts between participant and provider via content analysis of communication interactions (ie, information sought, reassurance given, and call was of a social nature). We also assessed whether the provider was checking progress reports and graphs and how the participant rated communication; this information was recorded on tracking forms and assessed at weekly team meetings. The goal for the provider was to use the *talk* model 90% of the time, and that 90% of participants would report high satisfaction. Acceptability was defined as an endorsement and measured by the number and types of problems encountered, such as difficulty using accelerometer and app, and satisfaction with the communication system. Implementation procedures included participant and wound care provider recommendations used to refine CALF or BEAT and the number and types of refinements made. Maintenance was defined as the number of patients who would continue the intervention, and the provider perception of impact and potential for future apps. Safety was evaluated by recording any adverse effects or safety issues such as cramping or new leg pain that occurred during the study period. Treatment fidelity was monitored weekly and assessed retrospectively from data obtained from BEAT in terms of the number of daily exercises completed per participant over the length of the study.

### Statistical Analysis

The study sample was characterized using descriptive statistics. Primary continuous feasibility outcome measures were adherence and acceptability and patient-physician interactions. Additional quantitative measures of feasibility included recruitment and dropout proportions and patient-provider communication measures (ie, clarity of reports, endorsement, potential for adoption, usability of BEAT, and mobile phone). For the continuous measures such as total number completed daily CALF exercises per level, means and medians were obtained. Owing to the small sample size, normally distributed data were not assumed. No hypothesis testing was carried out, and therefore, no *P* values are provided in concordance with recommendations in the CONSORT 2010 statement for randomized, pilot and feasibility trials [24]. Similarly, no effect sizes (eg, Cohen *d*) were provided because of large imprecision with small sample sizes [25]. Instead, 95% CIs for differences in medians between groups were obtained using quantile regression, whereas differences in proportions of categorical feasibility outcomes with their corresponding 95% CIs were obtained using Newcombe risk difference method, to describe estimates of the magnitude of the clinical effects. The technology (problems and acceptability) was evaluated through the documentation on tracking forms and iterative processes using participant interviews. All data were entered on a password-protected Web-based data management system Research Electronic Data Capture and analyzed using SAS software version 9.4 (2016; SAS Institute Inc, Cary, North Carolina).

## Results

### Demographic Data

Age, sex, comorbid conditions, medications, and residence (rural/urban) and other population characteristics are shown in Table 1.

The sample was predominantly white, female, and obese with an average ulcer age of greater than 4 months duration and ulcers occurred most frequently on the lateral and medial aspects of the lower extremities. Several differences were noted in age, employment, education, and marital status, but because of the small sample size, these differences were not assessed for significance or their relationship with feasibility outcomes.

**Table 1.** Demographic and clinical characteristics by FOOTFIT and FOOTFIT+ groups at baseline (N=24).

| Characteristic | FOOTFIT (n=12) | FOOTFIT+ (n=12) |
|---|---|---|
| Age (years), mean (SD) | 60.7 (13.7) | 69.1 (11.5) |
| **Sex, n (%)** | | |
| Females | 8 (67) | 6 (50) |
| **Race, n (%)** | | |
| Black/African American | 4 (33) | 6 (50) |
| White | 8 (67) | 6 (50) |
| **Ethnicity, n (%)** | | |
| Non-Hispanic/Latino | 12 (100) | 12 (100) |
| **Marital status, n (%)** | | |
| Never married | 2 (17) | 3 (25) |
| Married | 7 (58) | 3 (25) |
| Separated | 0 | 2 (17) |
| Divorced | 2 (17) | 1 (8) |
| Widowed | 1 (8) | 3 (25) |
| **Educational level, n (%)** | | |
| Eighth grade or less | 0 | 1 (8) |
| Some high school | 2 (17) | 1 (8) |
| High school graduate | 2 (17) | 3 (26) |
| Some college | 5 (41) | 2 (17) |
| College graduate | 2 (17) | 5 (41) |
| Post graduate and/or higher level degree | 1 (8) | 0 |
| **Employment, n (%)** | | |
| Employed fulltime | 1 (8) | 2 (17) |
| Not employed | 5 (42) | 1 (8) |
| Retired | 6 (50) | 9 (75) |
| **Job classification, n (%)** | | |
| Professional | 4 (33) | 5 (42) |
| Technical | 1 (8) | 2 (16) |
| Manual | 7 (59) | 5 (42) |
| **Residence, n (%)** | | |
| Urban | 5 (42) | 9 (75) |
| Rural | 7 (58) | 3 (25) |
| Weight (lb), mean (SD) | 292.5 (104.0) | 235.0 (58.9) |
| BMI ($kg/m^2$), mean (SD) | 45.2 (14.5) | 35.5 (8.8) |
| **VLU[a] location, n (%)** | | |
| Proximal | 1 (8) | 1 (8) |
| Distal | 4 (33) | 4 (33) |
| Medial | 6 (50) | 3 (25) |
| Lateral | 9 (75) | 8 (67) |
| Anterior | 4 (33) | 2 (17) |
| Posterior | 0 | 2 (17) |

| Characteristic | FOOTFIT (n=12) | FOOTFIT+ (n=12) |
|---|---|---|
| Recurrent VLU (yes), n (%) | 10 (83) | 5 (58) |
| Age of ulcer (days), mean (SD) | 1050.7 (1101.4) | 813.8 (928.8) |
| Age of ulcer (months), mean (SD) | 35.0 (36.7) | 27.1 (31.0) |
| Number of ulcers, mean (SD) | 7.2 (4.9) | 5.2 (9.8) |
| **Comorbid conditions (top 5), n (%)** | | |
| Hypertension | 9 (75) | 7 (58) |
| Arthritis | 8 (66) | 6 (50) |
| Diabetes | 4 (33) | 6 (50) |
| Thyroid problems | 6 (50) | 2 (17) |
| Varicose veins | 2 (17) | 4 (33) |
| Vein stripping | 4 (33) | 2 (17) |
| **Medications (top 5), n (%)** | | |
| Antihypertensive med | 9 (75) | 6 (50) |
| Cholesterol | 6 (50) | 6 (50) |
| Pain pills | 7 (58) | 3 (25) |
| Diabetes pills | 4 (33) | 5 (42) |
| Diuretics | 4 (33) | 4 (33) |

[a]VLU: venous leg ulcer.

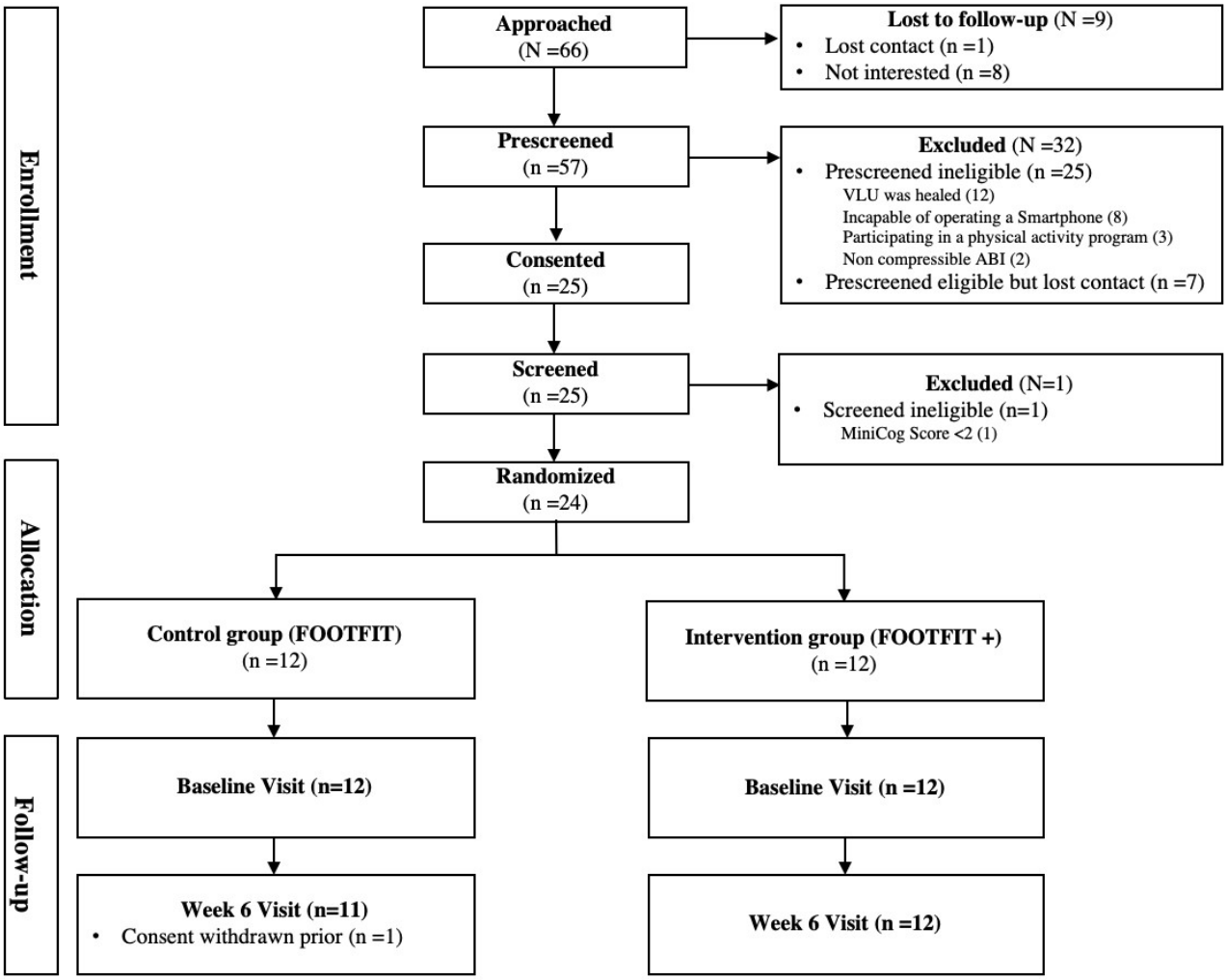
## Feasibility Outcomes

### *Reach*

Reach data are shown in Figure 1 as the number of patients approached, prescreened eligible, consented, screened, and enrolled.

The sample was representative of the target population except for lacking one Hispanic/non-white participant. Although flyers were posted around clinics and presentations were made by study staff at various venues such as senior living residences and health fairs, all participating in the study were recruited by referral of wound providers in clinic settings. One recruitment goal was to consent and enroll 1 out of 5 patients approached; however, we were unable to track this because of providers' lack of consistent documentation on study logs. There were no reported adverse events associated with the intervention.

**Figure 1.** Participant Consolidated Standards of Research Trials flowchart.




### *Adoption*

Patient-provider communication was the focus of adoption; however, none of the participants contacted their provider during the study. Similarly, none of the 3 providers reviewed progress reports and graphs citing time as the major limitation.

### *Acceptability*

Acceptability was considered low. The participants encountered several problems with BEAT, the phone, and the app including difficulty remembering to turn BEAT and the phone on and off before and after performing CALF; most participants forgot to close the app at the end of the day, which was needed to reset it to capture data from the next session. The button on BEAT was small and almost flush with the device, making it hard for participants to see and feel to turn it on and off. The phone screen (iPhone 4) was small and challenging for individuals to view. Many participants reported that having to double click the home screen button to close the app was tricky because of arthritic fingers or long fingernails. The app did not work well when individuals wanted to replay the introduction video, and despite attempts by the programmer, this problem could not be resolved. Owing to participants' ulcer treatments with multilayer compression wraps applied from the toes to the knee, the slippers on which BEAT was affixed did not fit well, and the device sometimes popped off during use. Participants reported that they were *annoyed* by the overly frequent (at a minimum 3 times per day) reminder text or email alerts to start the exercises. If participants were unable to do the foot exercises at that time, the reminder would continue to sound. We observed that although participants demonstrated proficiency during return demonstration of using the accelerometer and app during the baseline visit, several participants became confused when they returned home, prompting calls to study staff for questions although instructions were available in the app.

### *Implementation*

Refinements to implementation were mainly related to the app. The app was designed for iOS 8 or later, developed on Heroku cloud platform, and beta tested through Apple TestFlight. Every 3 months, the app's beta testing period expired, and a new build of the app had to be uploaded by the developers to TestFlight for actively using the app. At these times, the app also had to be updated and reinstalled, often leading to participant confusion and the inability to perform the leg exercises. The providers had no specific recommendations for refinements to reports of participant involvement other than they were too busy to review progress on a regular basis. In response to this challenge, study staff sent the reports each week to the provider, rather than the provider having to access the app site database to search and

review data for each participant. The providers did not communicate feedback *back* from the reports to participants. Providers reported they inconsistently reviewed these reports if at all.

### *Maintenance*

Participants indicated they liked FOOTFIT, but because of many technical glitches, they believed they would rather do the exercises on their own and not rely on the system. Providers reported that participants told them the exercises helped relieve stiffness and lower leg pain and *enjoyed being in the study*. However, they also revealed to the providers that it was often difficult to operate the accelerometer and phone, corroborating the information reported to study staff.

### *Adherence*

Duration and frequency of BEAT use to perform CALF is shown in Table 2. The duration was mostly similar between the two groups; however, adherence to the recommended frequency of use 3 times per day was lower in the FOOTFIT group (17%) than in the FOOTFIT+ group (50%).

**Table 2.** Distribution of adherence level to recommended exercise duration by FOOTFIT and FOOTFIT+ groups (N=24).

| Adherence | FOOTFIT (n=12) | FOOTFIT+ (n=12) |
|---|---|---|
| **Duration of exercising (≥15 seconds), n (%)** | | |
| 85%-89% | 1 (8) | 1 (8) |
| 90%-94% | 2 (17) | 0 (0) |
| 95%-99% | 5 (42) | 6 (50) |
| 100% | 4 (33) | 5 (42) |
| **Frequency, n (%)** | | |
| **Always adherent ( ≥ 85%)** | | |
| 85%-89% | 1 (8) | 1 (8) |
| 90%-94% | 0 (0) | 3 (25) |
| 95%-99% | 1 (8) | 1 (8) |
| 100% | 0 | 1 (8) |

### *Exercise Frequency*

During weeks 1 and 2 (CALF exercise level 1 intensity), across the two groups, 33% (8/24) of the participants were adherent (moved foot at or above recommended duration of ≥15 seconds per exercise) during all exercises at each session (Table 3).

During weeks 3 and 4 (CALF exercise level 2 intensity), 59% (13/22) of the total sample were always adherent with fewer (3/11, 27%) not always adhering in the FOOTFIT group compared with 55% (6/11) in the FOOTFIT+ group. Conversely, sample participants showed lower adherence during weeks 5 and 6 (CALF exercise level 3 intensity) with an overall 53% (10/19) of the participants always adherent. Almost 50% more individuals in the FOOTFIT group (6/10, 60%) were not always adherent compared with the FOOTFIT+ group (3/9, 33%).

**Table 3.** Frequency distribution for adherence across all exercises by the level of intensity for FOOTFIT and FOOTFIT+ groups.

| Conditioning activities for lower leg function exercise intensity level[a] | FOOTFIT (n=12) | FOOTFIT+ (n=12) | Difference in proportions[b] | 95% CI |
|---|---|---|---|---|
| **Level 1, n (%)** | | | 0 | −0.4 to 0.4 |
| Always adherent | 4 (33) | 4 (33) | | |
| Not always adherent | 8 (67) | 8 (67) | | |
| **Level 2, n (%)** | | | 0.3 | −0.1 to 0.7 |
| Always adherent | 8 (73) | 5 (45) | | |
| Not always adherent | 3 (27) | 6 (55) | | |
| **Level 3, n (%)** | | | 0.3 | −0.2 to 0.7 |
| Always adherent | 4 (40) | 6 (67) | | |
| Not always adherent | 6 (60) | 3 (33) | | |

[a]Level 1: weeks 1+2, level 2: weeks 3+4, and level 3: weeks 5+6.

[b]Difference and 95% CIs obtained using the Newcombe risk difference method.

### *Exercise Duration*

Overall there was an increase of approximately 6 seconds in median exercise duration between level 1 and level 3 exercises for the FOOTFIT but not the FOOTFIT+ group (Table 4).

**Table 4.** Mean (SD) and median (range) exercise duration in seconds by exercise level for FOOTFIT and FOOTFIT+ groups.

| Conditioning activities for lower leg function exercise levels[a] | FOOTFIT | FOOTFIT+ | Difference in medians (SE)[b] | 95% CI |
|---|---|---|---|---|
| **Level 1 (weeks 1+2), (n=12;12)** | | | 1.9 (6.5) | −11.7 to 15.4 |
| Mean (SD) | 29.0 (9.8) | 31.7 (18.9) | | |
| Median (range) | 24.3 (18.5-46.0) | 25.9 (16.1-82.9) | | |
| **Level 2 (weeks 3+4), (n=11;11)** | | | 0.3 (6.6) | 13.5 to 14.0 |
| Mean (SD) | 30.3 (17.6) | 30.9 (20.7 | | |
| Median (range) | 21.7 (16.4-71.0) | 21.4 (17.3-84.7) | | |
| **Level 3 (Weeks 5+6), (n=10;9)** | | | 4.0 (13.2) | 23.3 to 31.2 |
| Mean (SD) | 34.5 (18.4) | 36.2 (25.7) | | |
| Median (range) | 30.6 (17.8-74.6) | 26.3 (19.8-99.1) | | |

[a]Level 1: weeks 1+2, level 2: weeks 3+4, and level 3: weeks 5+6.

[b]Difference in medians (SE) and 95% CIs obtained using quantile regression.

## Discussion

### Principal Findings

In our study, we investigated the feasibility of FOOTFIT that combined a foot-worn accelerometer BEAT with a mHealth app and evidence-based foot exercises CALF; in the FOOTFIT+ group, an additional patient-provider communication option (FOOTFIT+) was included. The intervention was developed to promote adherence to a 6-week progressive exercise program suited to the needs and preferences of minimally ambulatory older adults with VLUs. The results showed participants had problems using the accelerometer and app but were mostly adherent to the exercise protocol. Minimal gains were made in exercise intensity and duration in both groups. It is important to point out that the study focused on feasibility and was not powered to detect differences between groups.

Our one-on-one coaching and MI physical activity interventions found to be feasible in our previous studies guided the development of our mHealth foot-worn sensor and exercise app intervention. Exercise is recommended for patients with VLUs to enhance calf muscle pump function, which, in turn, improves lower extremity function and may aid in wound healing. Supervised and unsupervised programs that incorporate resistance, flexibility, and moderate-intensity aerobics can be safely performed by individuals with VLUs [8]. Other benefits of exercise include improved microvascular circulation of the lower limb [26,27] and improved wound healing [9,10]; however, a recent review of six randomized controlled trials to examine the effects of exercise on healing showed high-quality evidence is lacking to support healing [6]. Many of the studies reviewed targeted fairly high-functioning adults residing in the community; had multiple intervention types such as 10,000 steps, behavior modification, or community-based groups; included small sample sizes; and measured numerous variables with disparate measures, making comparisons to make and draw conclusions across studies. Our feasibility studies of the use of mHealth devices specifically targeted functionally impaired, minimally ambulatory individuals; tracked and provided feedback on progress; and delivered motivational messages and reminders to enhance adherence to exercise.

### Adherence

Many of the participants in our study had little or no previous experience with accelerometers or apps; thus, we anticipated lower adherence rates, given the difficulty with using the device and phone app. However, 100% of the participants were adherent to our 6-week exercise program more than 85% of the time in terms of duration, whereas 33% of the participants adhered more than 85% in terms of frequency of exercises. This finding is lower than findings of a 12-week home-based program in which 59% of individuals adhered to exercise more than 75% of the time [8]. Current evidence of mHealth interventions to promote physical activity demonstrate positive short-term effects (increased daily step counts and minutes spent on physical activity); however, evidence for long-term effects is lacking [28]. We recognize our results may have differed because of the short intervention period and may not have captured decreased sustainability over time. We also point out that this study differs from our previous studies in delivery modalities (face-to-face interaction with wound provider during a clinic visit or one-on-one visit with a coach via social media visit). It would be interesting to determine which modality, including the use of an app, is preferable/acceptable to patients and the barriers and facilitators to adoption of each approach.

Furthermore, at the time of accelerometer and app development, the system may not have been ideally designed for an older population. Although fitness and other exercise apps are developing at a rapid pace for waist and wrist-worn devices, to our knowledge, our accelerometer was the first triaxial method used to measure minute foot movements. Having to push a very small button on the accelerometer proved to be a challenge for the participants, many of whom had visual problems, difficulty with bending over, and toggling through the phone app. Many

forgot to turn off the device, causing miscommunication with the app and draining the battery. A major problem with using the phone was the participants' inability to quickly double click the home button to turn the app on and off.

App features such as reminders have been found to be essential for app engagement and are influenced by social demographics; health characteristics; intentions to change; and actual health behaviors, motivation, and support [29]. In a study of app use for physical activity, the findings suggested that older individuals, males, having a degree, or less than high school education are associated with reduced likelihood of adopting health apps [30]. The findings from a study of health apps in older adults demonstrated increased levels of physical activity compared with those without a device or health app. It remains unclear whether mHealth interventions have a greater impact on adherence to physical activity recommendations than non-mHealth interventions such as printouts.

### Provider Engagement

We anticipated greater user engagement of the patient-provider communication feature. Participants in the FOOTFIT+ group were aware that they could communicate via text, emails, or phone with the provider, and perhaps, this knowledge improved their adherence rates from 33% always adherent to CALF level 1 to 67% to CALF level 3 compared with FOOTFIT 33% and 40%, respectively. The reasons expressed by participants for not calling the wound provider included "I didn't want to bother her," "I could call the study coordinator if I had a question or problem," and "I didn't have any problems I needed to tell my doctor and if I did, I talked to her during my wound clinic visit." The wound providers reported they did not have time to review the reports (these were compiled weekly by the study coordinator and sent directly via email to the providers) and noted they did not see the need to call participants as they thought they were doing well. In addition, the wound providers saw the participants weekly in their clinics and asked at that time how they were doing with the study. Thus, the patient-provider communication feature was viewed as not being necessary. However, in a longer study and without weekly clinic visits, this might have been a more useful option for patients to check in with their providers.

### Limitations and Strengths

The limitations of this randomized controlled trial include a small sample size, as the primary aim was feasibility study. In addition, because the accelerometer and app were specifically designed for individuals with VLUs, the generalizability of results is limited. However, other minimally ambulatory populations with chronic conditions that have limited function may benefit from this type of physical activity approach. We recognize the intervention did not include a specific adherence-enhancing component other than daily alerts; there were no provisions for behavioral change support. Behavioral approaches should be incorporated into future designs to enhance motivation and user engagement. In addition, our evidence-based CALF program was specifically designed for a progressive, short initial exercise *boost* for this minimally ambulatory population, prior to them undertaking a more rigorous physical activity program. Thus, our findings do not add to the fields of physical activity or exercise sciences in a way that advances our understanding of the influence of exercise on wound healing.

The strengths of the study include the use of a foot-worn device specifically designed for the population and the use of evidence-based exercises tailored to enhance foot function. The use of the RE-AIM as the feasibility framework provided both quantitative (number of log-ins) and qualitative (open-ended questions about usability) information and allowed for detailed participant perspectives and experiences using the intervention components. This mHealth study is unique to the best of our knowledge; it is the first study to focus on a minimally physically functional group of older adults to promote an exercise intervention focused solely on the foot and ankle.

## Conclusions

In this study, physically low-functioning adults with VLUs perceived the foot-based accelerometer BEAT, CALF exercises, and phone-based app (FOOTFIT) as somewhat acceptable and the added communication feature (FOOTFIT+) as not particularly useful. Most participants said they would perform the exercises without the accelerometers and reminders. As readily available apps continue to evolve and promote physical activity in populations with chronic conditions, our stand-alone intervention did not seem sufficient to promote exercise in this population. Future mHealth interventions should consider adding tailored adherence-enhancing components, such as added support (although the participants had access to the wound clinician) to positively influence behavior change, improve the functionality of the wearable device, and revise the app to make it more intuitive and bigger/easier to read on the phone and tablet. From our experience with three small trials, we advocate that providers such as wound specialists or primary care providers discuss patient preferences for engaging in the frequency and types of physical activity to enhance lower leg physical functioning.

## Acknowledgments

This project was supported by the National Institutes of Health, National Institute of Nursing Research (TK, primary principal investigator), #R21NR015134, and in part, by the National Center for Advancing Translational Sciences of the National Institutes of Health under Grant Number UL1 TR001450. 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

All authors were involved in the study design. MP led the data collection. MM performed the data analyses. MM provided project oversight. AV and AS designed, manufactured, and developed the accelerometer and app. TK was the principal investigator and drafted the manuscript. All authors critically reviewed and revised the manuscript. All authors read and approved the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT EHEALTH checklist (v 1.6.1).
[PDF File (Adobe PDF File), 375 KB - mhealth_v8i5e15015_app1.pdf ]

## References


1. Abelyan G, Abrahamyan L, Yenokyan G. A case-control study of risk factors of chronic venous ulceration in patients with varicose veins. Phlebology 2018 Feb;33(1):60-67. [doi: 10.1177/0268355516687677] [Medline: 28077026]
2. O'Brien J, Finlayson K, Kerr G, Edwards H. The perspectives of adults with venous leg ulcers on exercise: an exploratory study. J Wound Care 2014 Oct;23(10):496-8, 500. [doi: 10.12968/jowc.2014.23.10.496] [Medline: 25296351]
3. Parker CN, Finlayson KJ, Edwards HE. Predicting the likelihood of delayed venous leg ulcer healing and recurrence: development and reliability testing of risk assessment tools. Ostomy Wound Manage 2017 Oct;63(10):16-33 [FREE Full text] [Medline: 29091035]
4. Phillips P, Lumley E, Duncan R, Aber A, Woods HB, Jones GL, et al. A systematic review of qualitative research into people's experiences of living with venous leg ulcers. J Adv Nurs 2018 Mar;74(3):550-563. [doi: 10.1111/jan.13465] [Medline: 28960514]
5. Williams KJ, Ayekoloye O, Moore HM, Davies AH. The calf muscle pump revisited. J Vasc Surg Venous Lymphat Disord 2014 Jul;2(3):329-334. [doi: 10.1016/j.jvsv.2013.10.053] [Medline: 26993396]
6. Smith D, Lane R, McGinnes R, O'Brien J, Johnston R, Bugeja L, et al. What is the effect of exercise on wound healing in patients with venous leg ulcers? A systematic review. Int Wound J 2018 Jun;15(3):441-453. [doi: 10.1111/iwj.12885] [Medline: 29446252]
7. Kirsner RS. Exercise for leg ulcers: 'Working Out' the nature of venous ulcers. JAMA Dermatol 2018 Nov 1;154(11):1257-1259. [doi: 10.1001/jamadermatol.2018.2926] [Medline: 30285045]
8. Klonizakis M, Tew GA, Gumber A, Crank H, King B, Middleton G, et al. Supervised exercise training as an adjunct therapy for venous leg ulcers: a randomized controlled feasibility trial. Br J Dermatol 2018 May;178(5):1072-1082 [FREE Full text] [doi: 10.1111/bjd.16089] [Medline: 29077990]
9. O'Brien J, Finlayson K, Kerr G, Edwards H. Evaluating the effectiveness of a self-management exercise intervention on wound healing, functional ability and health-related quality of life outcomes in adults with venous leg ulcers: a randomised controlled trial. Int Wound J 2017 Feb;14(1):130-137. [doi: 10.1111/iwj.12571] [Medline: 26817648]
10. Mutlak O, Aslam M, Standfield N. The influence of exercise on ulcer healing in patients with chronic venous insufficiency. Int Angiol 2018 Apr;37(2):160-168 [FREE Full text] [doi: 10.23736/S0392-9590.18.03950-0] [Medline: 29368880]
11. Kelechi TJ, Green A, Dumas B, Brotherton SS. Online coaching for a lower limb physical activity program for individuals at home with a history of venous ulcers. Home Healthc Nurse 2010;28(10):596-605. [doi: 10.1097/NHH.0b013e3181f7e99d] [Medline: 21057229]
12. Kelechi TJ, Mueller M, Spencer C, Rinard B, Loftis G. The effect of a nurse-directed intervention to reduce pain and improve behavioral and physical outcomes in patients with critically colonized/infected chronic leg ulcers. J Wound Ostomy Continence Nurs 2014;41(2):111-121. [doi: 10.1097/WON.0000000000000009] [Medline: 24413659]
13. Vorrink SN, Kort HS, Troosters T, Lammers JW. A mobile phone app to stimulate daily physical activity in patients with chronic obstructive pulmonary disease: development, feasibility, and pilot studies. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e11 [FREE Full text] [doi: 10.2196/mhealth.4741] [Medline: 26813682]
14. Plow M, Golding M. Using mHealth technology in a self-management intervention to promote physical activity among adults with chronic disabling conditions: randomized controlled trial. JMIR Mhealth Uhealth 2017 Dec 1;5(12):e185 [FREE Full text] [doi: 10.2196/mhealth.6394] [Medline: 29196279]
15. Estabrooks PA. Physical activity promotion and translational research. Transl Behav Med 2017 Mar;7(1):1-2 [FREE Full text] [doi: 10.1007/s13142-016-0456-y] [Medline: 28032316]
16. Ainsworth BE, Caspersen CJ, Matthews CE, Mâsse LC, Baranowski T, Zhu W. Recommendations to improve the accuracy of estimates of physical activity derived from self report. J Phys Act Health 2012 Jan;9(Suppl 1):S76-S84 [FREE Full text] [doi: 10.1123/jpah.9.s1.s76] [Medline: 22287451]

17. van Hecke A, Verhaeghe S, Grypdonck M, Beele H, Flour M, Defloor T. Systematic development and validation of a nursing intervention: the case of lifestyle adherence promotion in patients with leg ulcers. J Adv Nurs 2011 Mar;67(3):662-676. [doi: 10.1111/j.1365-2648.2010.05539.x] [Medline: 21198805]
18. Shaporev A, Gregoski M, Reukov V, Kelechi T, Kwartowitz DM, Treiber F, et al. Bluetooth enabled acceleration tracking (BEAT) mHealth system: validation and proof of concept for real-time monitoring of physical activity. E-Health Telecommun Syst Netw 2013;2(3):49-57. [doi: 10.4236/etsn.2013.23007]
19. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. Stud Health Technol Inform 2013;192:657-661. [Medline: 23920638]
20. Borson S, Scanlan JM, Chen P, Ganguli M. The Mini-Cog as a screen for dementia: validation in a population-based sample. J Am Geriatr Soc 2003 Oct;51(10):1451-1454. [doi: 10.1046/j.1532-5415.2003.51465.x] [Medline: 14511167]
21. Elwyn G, Tsulukidze M, Edwards A, Légaré F, Newcombe R. Using a 'talk' model of shared decision making to propose an observation-based measure: Observer OPTION 5 Item. Patient Educ Couns 2013 Nov;93(2):265-271. [doi: 10.1016/j.pec.2013.08.005] [Medline: 24029581]
22. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
23. Glasgow RE, Dickinson P, Fisher L, Christiansen S, Toobert DJ, Bender BG, et al. Use of RE-AIM to develop a multi-media facilitation tool for the patient-centered medical home. Implement Sci 2011 Oct 21;6:118 [FREE Full text] [doi: 10.1186/1748-5908-6-118] [Medline: 22017791]
24. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. Br Med J 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
25. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
26. Tew GA, Gumber A, McIntosh E, Kesterton S, King B, Michaels JA, et al. Effects of supervised exercise training on lower-limb cutaneous microvascular reactivity in adults with venous ulcers. Eur J Appl Physiol 2018 Feb;118(2):321-329 [FREE Full text] [doi: 10.1007/s00421-017-3772-0] [Medline: 29197931]
27. O'Brien J, Finlayson K, Kerr G, Shortridge-Baggett L, Edwards H. Using a theoretical approach to identify factors influencing adherence to an exercise programme for adults with venous leg ulcers. J Health Psychol 2018 Apr;23(5):691-700. [doi: 10.1177/1359105316656241] [Medline: 27387513]
28. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. Maturitas 2018 Jul;113:32-39. [doi: 10.1016/j.maturitas.2018.04.010] [Medline: 29903646]
29. 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 5;16(Suppl 5):290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
30. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]


## Abbreviations

**BEAT:** Bluetooth-enabled acceleration tracking
**CALF:** conditioning activities for lower leg function
**CONSORT:** Consolidated Standards of Research Trials
**mHealth:** mobile health
**MI:** motivational interviewing
**RE-AIM:** reach, effectiveness, adoption, implementation, and maintenance
**VLU:** venous leg ulcers


*Edited by G Eysenbach; submitted 12.06.19; peer-reviewed by G Signorelli, K Goessler, T Møller; comments to author 03.10.19; revised version received 04.12.19; accepted 02.02.20; published 15.05.20.*

*Please cite as:*
*Kelechi TJ, Prentice MA, Mueller M, Madisetti M, Vertegel A*
*A Lower Leg Physical Activity Intervention for Individuals With Chronic Venous Leg Ulcers: Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2020;8(5):e15015*
*URL: https://mhealth.jmir.org/2020/5/e15015*
*doi:10.2196/15015*
*PMID:32412419*

Original Paper

# Selecting Evidence-Based Content for Inclusion in Self-Management Apps for Pressure Injuries in Individuals With Spinal Cord Injury: Participatory Design Study

Maddalena Fiordelli[1,2], PhD; Claudia Zanini[2,3], PhD; Julia Amann[2,4], PhD; Anke Scheel-Sailer[3,5], Dr med; Mirjam Brach[2,3], MBA, MPA; Gerold Stucki[2,3], Prof Dr, Dr med; Sara Rubinelli[2,3], Prof Dr

[1]Institute of Communication and Health, Faculty of Communication Science, Università della Svizzera italiana, Lugano, Switzerland

[2]Swiss Paraplegic Research, Nottwil, Switzerland

[3]Department of Health Sciences and Medicine, University of Lucerne, Lucerne, Switzerland

[4]Department of Health Sciences and Technology, Health Ethics and Policy Lab, ETH Zürich, Zürich, Switzerland

[5]Swiss Paraplegic Centre, Nottwil, Switzerland

**Corresponding Author:**
Maddalena Fiordelli, PhD
Institute of Communication and Health
Faculty of Communication Science
Università della Svizzera italiana
Via Buffi 13
Lugano, 6900
Switzerland
Phone: 41 58 666 47 57
Email: maddalena.fiordelli@usi.ch

## Abstract

**Background:** Technological solutions, particularly mobile health (mHealth), have been shown to be potentially viable approaches for sustaining individuals' self-management of chronic health conditions. Theory-based interventions are more successful, as evidence-based information is an essential prerequisite for appropriate self-management. However, several reviews have shown that many existing mobile apps fail to be either theoretically grounded or based on evidence. Although some authors have attempted to address these two issues by focusing on the design and development processes of apps, concrete efforts to systematically select evidence-based content are scant.

**Objective:** The objective of this study was to present a procedure for the participatory identification of evidence-based content to ground the development of a self-management app.

**Methods:** To illustrate the procedure, we focused on the prevention and management of pressure injuries (PIs) in individuals with spinal cord injury (SCI). The procedure involves the following three steps: (1) identification of existing evidence through review and synthesis of existing recommendations on the prevention and self-management of PIs in SCI; (2) a consensus meeting with experts from the field of SCI and individuals with SCI to select the recommendations that are relevant and applicable to community-dwelling individuals in their daily lives; and (3) consolidation of the results of the study.

**Results:** In this case study, at the end of the three-step procedure, the content for an mHealth intervention was selected in the form of 98 recommendations.

**Conclusions:** This study describes a procedure for the participatory identification and selection of disease-specific evidence and professional best practices to inform self-management interventions. This procedure might be especially useful in cases of complex chronic health conditions, as every recommendation in these cases needs to be evaluated and considered in light of all other self-management requirements. Hence, the agreement of experts and affected individuals is essential to ensure the selection of evidence-based content that is considered to be relevant and applicable.

***(JMIR Mhealth Uhealth 2020;8(5):e15818)*** doi:10.2196/15818

**KEYWORDS**

mHealth; paraplegia; tetraplegia; pressure ulcers; consensus meeting; community engagement; recommendations

## Introduction

### Background

Ever since communication technologies were adopted for health care purposes and defined under the umbrella term electronic health (eHealth), the concept of empowerment and the use of technological solutions have become intertwined [1]. As technological devices became more personal and connected, this relationship took on a new relevance. In particular, mobile health (mHealth) solutions, commonly defined as the use of mobile and wireless technologies to support the achievement of health objectives [2], have been used to enhance the self-management of various chronic conditions [3], such as diabetes [4] and asthma [5]. Evidence indicates that these mHealth solutions can foster self-management by addressing multiple risk factors [6] and sustaining long-term adherence to prevention measures, which remains a major issue [7].

Studies have examined not only the effectiveness of mHealth [8], but also its design and development process. Even though there is great potential for using mobile technologies for health purposes, findings show that many of the existing mobile apps are not theoretically grounded [9] and their contents are not based on evidence [3,7,9,10]. This is problematic because studies mentioned that theory-driven health interventions are more effective than those without theoretical grounding [11]. It is only recently that mHealth has started to adopt strategies informed by behavior change theories, but this adoption has not been systematic [12]. Some apps are only partially applying the principles of behavior change theories [13], whereas others have defined the app's features or its mechanism (ie, goal setting) based on a set of different theories or models, but without clear reference to them [14-16]. Additionally, some authors have attempted to create a framework to develop digital behavior change interventions that integrate, for instance, behavior theories, design thinking, and user-centered design [17,18]. Despite these efforts, to date, many apps are still not based on theory, as attested by recent systematic reviews [19,20].

Another flaw of many existing apps lies in the quality of their content, which does not reflect the latest scientific evidence. Indeed, some ex-post examinations of mHealth apps underlined that their content rarely adheres to evidence-based knowledge [21-24]. Some authors have based the content of their interventions on the results of systematic reviews or additional participatory efforts (ie, involving different stakeholders) [25,26]. However, their procedures are not detailed and cannot be replicated. So far, the efforts to develop a framework for integrating evidence-based content into mHealth interventions have been limited [27]. Hence, it remains unclear how disease-specific recommendations and professional best practices should be selected to inform mHealth interventions. This is problematic as evidence-based information can enhance health literacy [28], which is a precondition for patient participation and informed decision-making [29]. Consequently, apps that are based on outdated or inaccurate content might negatively affect the users' health and safety [30-32]. Considering the huge amount of incorrect and misleading information available on the internet, as well as in leaflets and other lay publications [33,34], it is of utmost importance that new mHealth interventions tackle the issue of content quality.

Participatory design is a democratic process involving different stakeholders from the early phases of the design process [35-37]. At least the following two premises provide the basis for different participatory design approaches: all stakeholders should be involved in the design phase to inform the approach and this will increase the likelihood of technology acceptance because it will help set clear expectations [38]. It is for a very good reason that many authors underscored the potential of a participatory design approach throughout various steps, such as requirement analysis, definition of features, and user interface design [39-41], but without providing much clarity on the most appropriate involvement of experts and other stakeholders for the selection of content. Participatory design could be a viable approach for achieving the evidence basis of an app. Several guidelines exist, but they are mostly designed for health care professionals rather than for community-dwelling individuals or patients. Selecting the content of an app through a participatory design approach involves understanding which of the recommendations are not only impactful in terms of prevention, but also feasible and applicable for people living in the community.

The objective of this study was to fill this gap by describing a structured procedure for the participatory identification of evidence-based content to ground the development of a self-management app. To illustrate the procedure, we used a project based in Switzerland aiming to develop an app for the prevention and self-management of pressure injuries (PIs) in individuals with spinal cord injury (SCI).

### A Case in Point: Spinal Cord Injury

SCI is a complex chronic condition affecting human functioning in all aspects [42], and it is associated with a number of complications [43,44]. People with SCI have a high risk of developing PIs [42]. The incidence of PIs in the SCI population is 25% to 66% [45], and approximately 85% of individuals with SCI will experience PIs at some point in their lifetimes [46]. PIs impact the quality of life of the affected individuals, as their treatment necessitates prolonged inactivity, which often results in a loss of income and a feeling of social isolation [47]. Moreover, evidence shows that PIs can account for approximately one-fourth of the cost of care for individuals with SCI [48].

There is general agreement on the fact that PIs might be often preventable in individuals with SCI [49] and that prevention is more cost-effective than treatment [50]. Prevention is possible through active self-management. However, this self-management remains challenging owing to the many different factors that need to be taken into account. Indeed, individuals with SCI have to play an active role in the prevention of PIs by, for instance, adapting their behavior, which includes repositioning, performing pressure-relieving movements, and keeping the skin clean [51]. The prevention and management of PIs in individuals with SCI could benefit from the development of an evidence-based mobile app that supports individuals in performing the many preventive measures, as well as monitoring and treating early stage PIs.

## Methods

### Study Design

We used a consensus method for the participatory identification of evidence-based content to ground the development of a self-management app for PIs in individuals with SCI. Indeed, to ensure that individuals with SCI have access to sources of information that are credible, of good quality, and up to date, the information provided in the app should be consistent with the latest available clinical recommendations, including those that are indeed the best available evidence for pressure ulcer prevention and remain the foundation of a prevention program [52].

The recommendations were identified through a three-stage research procedure developed following the main steps of the consensus development method [53]. First, a review of existing recommendations for the prevention and management of PIs in individuals with SCI was conducted. Second, a consensus meeting [54] to select the most important recommendations that individuals with SCI should apply in their daily lives was performed. Finally, the results were consolidated by the expert team. Both the review of the recommendations and the consensus meeting were conducted at the end of 2017, while the third phase was performed during the first quarter of 2018.

### Stage 1: Review and Categorization of Existing Recommendations

Published recommendations on the prevention and management of PIs were identified through an electronic search and consultation with experts between March and July 2017. Keywords for the search were combined from three different domains. The first was related to PIs (ie, pressure ulcers, pressure injuries, decubitus, pressure sores, bedsores, and skin problems), the second was related to self-management (ie, prevention, detection, treatment, self-management, reduction, and risk factors), and the third was related to SCI (ie, spinal cord injury, tetraplegia, quadriplegia, and paraplegia). The online search applied these keywords in both English and German languages. The search was performed in Google as well as PubMed. The research team extracted all recommendations for the prevention and management of PIs that were directed toward or could be applied by community-dwelling individuals with SCI.

The review obtained a comprehensive collection of recommendations that were screened and synthesized (ie, similar recommendations from different sources were merged). The results of stage 1 were presented in a document that was sent to all participants of the consensus meeting for preparation.

### Stage 2: Consensus Meeting

A purposive sample of health professionals and community-dwelling individuals with SCI were invited to participate in a consensus meeting [53]. With the help of SCI-specialized medical doctors, we identified health professionals who may have experience with PIs in individuals with SCI. For the recruitment of those working in the inpatient setting, we contacted the different departments of the four SCI rehabilitation centers in Switzerland and requested for collaboration. For the recruitment of health professionals working in the outpatient setting and individuals with SCI, we relied on informal networks. Through this process we contacted a total of 35 individuals, and they were offered two possible dates for the consensus meeting.

The final sample of 15 participants [53] included SCI-specialized medical doctors, nurses, wound experts, psychologists, occupational therapists, physiotherapists, and nutritionists who were from different parts of Switzerland and working in SCI rehabilitation centers in Switzerland, as well as home care providers, home care counsellors, representatives from an accident insurance fund, and individuals with SCI. Table 1 presents the participants' characteristics.

The consensus meeting was grounded in a systematic consensus planning process that helps to prioritize issues of a different kind during experts' discussions [55]. The meeting lasted one day and was structured in two parts. A person experienced in consensus meetings moderated the plenary sessions. Three persons facilitated the working groups. They were trained for the technical tasks (eg, dealing with the voting system) and were knowledgeable about the project.

**Table 1.** Characteristics of participants in the consensus meeting.

| Role | Workplace (for HPs[a]) | Working experience as a HP (years)/Years as a wheelchair user | Year of birth | Gender |
|---|---|---|---|---|
| SCI[b]-specialized medical doctor | Swiss Paraplegic Centre | 15 | 1962 | M |
| SCI-specialized medical doctor | Clinique de réadaptation romande | 14 | 1972 | M |
| SCI-specialized nurse/wound expert | Swiss Paraplegic Centre | 18 | 1966 | F |
| SCI-specialized nurse/wound expert | Clinique de réadaptation romande | 16 | 1974 | M |
| SCI-specialized nurse/wound expert | Clinique de réadaptation romande | 2 | 1992 | F |
| SCI-specialized nurse/wound expert | SCI-specialized counseling service | 14 | 1973 | F |
| Occupational therapist | REHAB Basel | 16 | 1971 | F |
| Physiotherapist | Swiss Paraplegic Centre | 18 | 1965 | F |
| Nutritionist | Swiss Paraplegic Centre | 2 | 1991 | F |
| Psychologist | Swiss Paraplegic Centre | 7 | 1964 | F |
| Psychologist and person with SCI | Balgrist Klinik | 28/33 | 1957 | M |
| Home care provider | Home care service | 8 | 1990 | F |
| Home care counsellor | SCI-specialized counseling service | 3 | 1972 | F |
| Representative from an accident insurance fund | Swiss Accident Insurance Fund (Suva) | 8 | 1967 | M |
| Person with SCI | N/A[c] | N/A/27 | 1970 | M |

[a]HP: health professional.

[b]SCI: spinal cord injury.

[c]N/A: not applicable.

### *Consensus Meeting Part I: Recommendations Selection*

The participants were divided into two working groups (whenever possible, a representative for every profession and a person with SCI were included in each working group). Moreover, professionals who worked together were included in different groups. They were asked to discuss one by one the recommendations derived from stage 1 and to vote by show of hands in favor of or against their inclusion in the set of recommendations to be implemented in the app. The vote should be based on the relevance and applicability of the recommendations for community-dwelling individuals with SCI. The facilitator of each group was in charge of taking notes on the discussions and carrying out the vote with the help of an ad-hoc technological infrastructure. A Microsoft Access (2010, version 14.0; Microsoft Corp, Redmond, Washington, USA) database containing the list of recommendations resulting from stage 1 was developed prior to the consensus meeting. Every participant voted in favor or against inclusion of each of the recommendations. The facilitator entered the sum of individual votes into the Access database. Based on this sum, a percentage of agreement for including each recommendation was computed. After this first vote (vote A), it was possible to merge the votes of the working groups and retrieve from the system the list of recommendations divided into recommendations to be included, recommendations to be excluded, and recommendations that were ambiguous. As the consensus method is based on a democratic debate and judicial model [53], the recommendations voted on below 40% were excluded, the recommendations voted on above 75% were included, and the recommendations voted on between 40% and 75% were considered ambiguous. These thresholds have been defined based on the experience of previous consensus meetings. The last group of recommendations was discussed in a plenary session in which all participants could argue in favor of or against their choice. After this exchange, the working groups met again to vote on the ambiguous recommendations (vote B). The recommendations were included, excluded, or considered ambiguous following the same rules as in vote A.

Moreover, during the group discussions, the participants had the opportunity to indicate that a recommendation needed specification. This was mostly the case when the recommendation was deemed to be too generic or when its applicability for community-dwelling individuals with SCI was considered unclear or vague. The recommendations that needed specification were collected in a list and further elaborated on in an afternoon session (part II).

### *Consensus Meeting Part II: Recommendations Specification*

The participants were divided into three working groups that were stratified by profession, workplace, and affiliation with the previous working groups. As for the morning working groups, whenever possible, we distributed the participants so that at least one representative of every profession and of people with SCI was present in each group. Moreover, professionals who worked together were included in different groups. We also differently mixed the participants with respect to the morning working groups.

Participants further specified the recommendations that were indicated during the previous session as being too vague or unclear to be implemented by community-dwelling individuals with SCI in their daily lives. Each of the three working groups received a list of 20 or 21 recommendations to specify (total 62) and was assigned a sheet of paper presenting a research-based user persona. User personas (Multimedia Appendix 1), which are fictional characters with concrete characteristics and behaviors that are intended to represent different user types, have been used in the user-centered design process for designing software [56,57]. They helped make the specification process concrete, as each group could refer to a vivid portrait.

### Stage 3: Consolidation of Results

After the consensus meeting, the research team together with two experts from the project scientific advisory board consolidated the results by refining them and taking into consideration the input of the participants. For instance, special attention was devoted to the recommendations that remained ambiguous after stage 2. They were screened and sorted out by the research team based on eight logical rules for their inclusion or exclusion. The rules (Textbox 1) referred to the size of the discrepancy between the results of vote A and vote B, and between the two working groups. Additionally, new recommendations were developed for domains that, according to the participants, were insufficiently covered by the existing recommendations. The consolidation stage resulted in a newer and more complete set of recommendations. These recommendations were shared with all the participants of the consensus meeting via email. Feedback from the participants was collected and integrated.

**Textbox 1.** Logical rules for the inclusion or exclusion of ambiguous recommendations.

1. If the average of group 1 (G1) and group 2 (G2) in vote A is <40% and in vote B is >40% but <75%, exclude the recommendation.
2. If the average of G1 and G2 in vote A is <40% and in vote B is >75%, include the recommendation.
3. If the average of G1 and G2 in vote A is >75%, average of G1 and G2 in vote B is >40% but <75%, and decrement of vote A-vote B is ≥25%, exclude the recommendation.
4. If the average of G1 and G2 in vote A is >75%, average of G1 and G2 in vote B is >40% but <75%, and decrement of vote A-vote B is <25%, include the recommendation.
5. If the average of G1 and G2 in vote A is >75% and in vote B is >75%, include the recommendation.
6. If the average of G1 and G2 in vote A is >40% but <75% and in vote B is >75%, include the recommendation.
7. If the average of G1 and G2 in vote A is >40% but <75%, average of G1 and G2 in vote B is >40% but <75%, and increment of vote A-vote B is ≥25%, include the recommendation.
8. If the average of G1 and G2 in vote A is >40% but <75%, average of G1 and G2 in vote B is >40% but <75%, and increment of vote A-vote B is <25%, exclude the recommendation.

## Results

### Stage 1: Review and Categorization of Existing Recommendations

The sources presented in Table 2 have been identified, and their documents have been systematically reviewed [58-65]. The recommendations extracted from the documents were categorized by applying a deductive-inductive approach. At first, the recommendations were ordered according to the four categories defined by Keast et al (ie, appropriate support surfaces, regular repositioning of the patient, optimizing nutrition, and skin care) [66]. This categorization, however, was not exhaustive. We therefore started an inductive process by grouping together those recommendations that were not covered by the categories defined by Keast et al. We created new categories until all recommendations belonged to one category. We then revised the categories with the aim of reducing their number. We compared among each other the recommendations included in every category and with those included in other categories, and when possible, we merged the categories. Following this procedure, we reached the number of 12 categories. This procedure is similar to the basic rule of the constant comparative method often used in qualitative research, namely the comparison of a new incident with the previous incidents coded in the same category [67].

The result of the review and recommendation categorization was a list of 130 recommendations for the prevention and management of PIs by individuals with SCI organized in relation to the following topics: (1) Support surface (code A); (2) Repositioning (code B); (3) Nutrition (code C); (4) Skin care (code D); (5) Skin assessment (code E); (6) Exercising (code F); (7) Collaboration with health professionals or caregivers (code G); (8) Transfers (code H); (9) Clothing (code I); (10) Body function and structure (code J); (11) Personal factors (code K); and (12) General (code L). The orders of the categories and recommendations within a category do not reflect a priority order. The recommendations were then collected in a preparatory

document, which was sent to the participants prior to the consensus meeting.

**Table 2.** Documents reviewed for the identification of recommendations.

| Source | Document |
|---|---|
| Deutschsprachige Medizinische Gesellschaft für Paraplegie (DMGP) (http://www.dmgp.de/) | Querschnittspezifische Dekubitusbehandlung und-prävention (2017) [58]<br>Psychologische Aspekte in der Dekubitusprophylaxe (2012) [59] |
| Ontario Neurotrauma Foundation (ONF) (http://www.onf.org) | Canadian Best Practice Guidelines for the Prevention and Management of Pressure Ulcers in People with Spinal Cord Injury. A Resource Handbook for Clinicians (2013) [60]<br>Preventing and Treating Pressure Sores. A Guide for People with Spinal Cord Injury (2015) [61] |
| European Pressure Ulcer Advisory Panel (http://www.ePIap.org/); National Pressure Ulcer Advisory Panel (http://www.nPIap.org/); and Pan Pacific Pressure Injury Alliance (EPIAP-NPIAP-PPPIA) (http://www.internationalguideline.com/) | Prevention and Treatment of Pressure Ulcers: Quick Reference Guide (2014) [62] |
| Spinal Cord Injury Research Evidence (SCIRE) (www.scireproject.com) | Pressure Ulcers Following Spinal Cord Injury (2014) [63] |
| International Spinal Cord Society (ISCoS) (http://www.iscos.org.uk/) | Textbook on Comprehensive Management of Spinal Cord Injuries, chapter 48 (2015) [64] |
| Schweizer Paraplegiker-Zentrum (SPZ) (http://www.paraplegie.ch) | Patientenaufklärung, Druckstellen-Dekubitus - V1.0 [65] |

## Stage 2: Consensus Meeting

### *Consensus Meeting Part I: Recommendations Selection*

Figure 1 shows the results of vote A. From the original list of 130 recommendations, 15 were excluded and 60 were included in the final set of recommendations for implementation in the app. The remaining 55 recommendations, with votes ranging between 40% and 74%, fell into the category of “ambiguous” and were subject to a second vote (vote B).

Figure 2 shows the results of vote B. Of 55 recommendations, nine were excluded and 25 were included in the final set of recommendations for implementation in the app. Twenty-one recommendations again fell into the category of “ambiguous.” These recommendations were no longer discussed by the participants during the consensus meeting, but were later examined by the research team.

### *Consensus Meeting Part II: Recommendations Specification*

A total of 62 recommendations needed specification. The list was composed of recommendations indicated by the working groups as well as recommendations indicated *a priori* by the research team. During the specification phase, different solutions for further clarification of the recommendations were defined by the groups. Most of the recommendations were specified by adding a further explanation of the action to take or by referring to additional criteria for the correct implementation of the recommendation. Approximately one-quarter of the specifications referred to the need to combine complementary recommendations. Experts suggested that a few of the recommendations should be specified by having a dedicated information section about the topic in the app. The specifications were gathered by the research team and further used in the development of evidence-based content for the app.

**Figure 1.** Results from vote A.

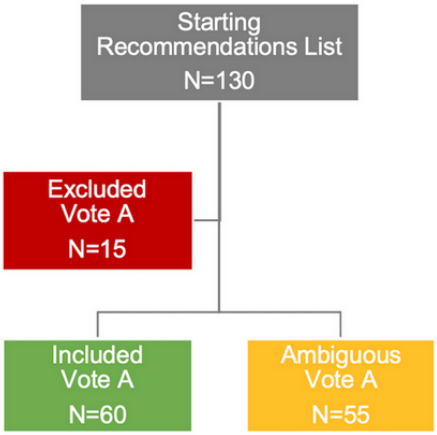
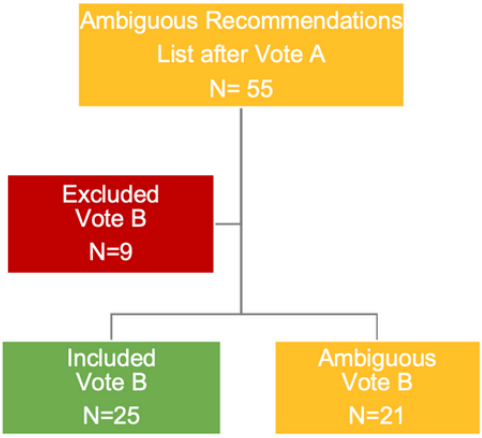


**Figure 2.** Results from vote B.




## Stage 3: Consolidation of Results

### *Decision on Ambiguous Recommendations*

The research team examined the 21 recommendations that remained ambiguous after vote B. The ex-post examination resulted in the inclusion of seven recommendations in the final set and the exclusion of 14.

### *Expert Consultations on the Category of Nutrition*

The review of existing recommendations (stage 1) resulted in three recommendations for the category of nutrition. These recommendations were debated considerably in the working groups during the consensus meeting (stage 2), as they were considered unsatisfactory. Although participants recognized the importance of nutrition as a risk factor for PIs, they criticized the incompleteness of the presented recommendations and their inability to depict the complexity of nutrition advice in relation to the prevention and management of PIs in individuals with SCI. Hence, the participants agreed with the research team to set up a working group composed of nutritionists and SCI-specialized medical doctors to develop new comprehensive nutrition recommendations. Table 3 provides details of the characteristics of the health professionals who were consulted to develop the recommendations about nutrition.

**Table 3.** Characteristics of the participants consulted for recommendations on nutrition.

| Role | Working place | Years of working experience | Year of birth | Gender |
|---|---|---|---|---|
| Nutritionist[a] | Swiss Paraplegic Centre (Nottwil) | 2 | 1991 | F |
| Nutritionist | Swiss Paraplegic Centre (Nottwil) | 6 | 1980 | F |
| Nutritionist | Swiss Paraplegic Centre (Nottwil) | 13 | 1976 | F |
| SCI[b]-specialized medical doctor | Swiss Paraplegic Centre (Nottwil) | 21 | 1966 | F |

[a]Took part also in the consensus meeting.

[b]SCI: spinal cord injury.

Multiple sessions of expert consultation were conducted with the aim of developing nutrition recommendations that account for the complex interaction between SCI management and PI prevention. These consultations took place between the end of 2017 and the middle of 2018. Based on previously examined and new sources [62,68-74], the nutritionists developed a new set of recommendations that encompassed information on drinking, weight, and nutrition. The proposed recommendations were then discussed and finalized in collaboration with SCI-specialized medical doctors. In total, six new nutrition recommendations were developed. They were circulated among the participants in the consensus meeting before being added to the final set of recommendations.

The final set of recommendations is presented in Multimedia Appendix 2. It includes 98 recommendations that synthesize evidence-based recommendations for the prevention and management of PIs for community-dwelling individuals with SCI.

A checklist for the process of participatory selection of the evidence to ground a self-management app is presented in Textbox 2.

**Textbox 2.** Checklist of the process of participatory selection of the evidence to ground a self-management app.

**Step 1: Systematic review and categorization of existing evidence**

- Theoretical
  - Broadly consider the concept of self-management for the respective health condition
  - Adopt a holistic perspective for care (bio-psycho-social)
- Methodological
  - Consult experts to identify the relevant sources for clinical guidelines
  - Identify and screen grey, scientific, and practice-oriented literature
  - Expand the search to different languages if possible
  - Synthetize the evidence (ie, merge and categorize it). The use of existing classifications might be of help.
- Practical
  - Draft a concise document summarizing the available evidence to be delivered to the participants of step 2 for preparation

**Step 2: Consensus meeting**

- Theoretical
  - Use an interdisciplinary approach
  - Adopt a holistic perspective for care (bio-psycho-social)
  - Review models of consensus building and strategies of conflict resolution
- Methodological
  - Identify relevant experts for participation in the consensus meeting
  - Balance the mix of experts in the consensus meeting (eg, in terms of position, years of working experience, age, and gender)
  - Provide the participants with user personas (fictional characters with concrete characteristics and behaviors that are intended to represent different user types)
  - Systematize the process of evidence selection to reach democratic decisions
  - Be attentive to potential gaps in current evidence pointed out by participants
- Practical
  - Facilitate participants' preparation for the meeting by delivering a concise document synthetizing the evidence and explaining the process
  - Train the moderator and facilitators of the working groups in advance (eg, to ensure that all participants express their opinion)
  - Provide support with a technological infrastructure to facilitate the voting and the calculation of the vote results
  - Allocate enough time for the different tasks and plenary discussions

**Step 3: Consolidation of results**

- Theoretical
  - Use an interdisciplinary approach
  - Adopt a holistic perspective to care (bio-psycho-social)
- Methodological
  - Compare and contrast the results with existing recommendations to identify potential gaps or "blind spots"
  - If needed, organize ad-hoc expert consultations
- Practical
  - Prepare a report explaining how results were reached
  - Elicit participant validation

## Discussion

### Principal Findings

This article proposes a procedure for the participatory identification of evidence-based content to ground the development of a self-management app. To our knowledge, this is one of the first attempts to apply a structured procedure for the participatory identification of evidence-based content for a self-management app in the field of SCI. The procedure consists of the following three steps: review of the literature, consensus meeting, and consolidation of the results (including, for instance, a set of expert consultations, if needed).

Our methodological approach raises two challenges that can hinder the development of evidence-based mHealth interventions. First, it has to be noted that sometimes the literature itself presents contradictory evidence [52,75,76], as the field of medicine is in continuous evolution. This underscores the challenge for clinicians and app developers in terms of identifying evidence-based knowledge on a topic. Thus, the involvement of experienced health care professionals might be a valuable means to assess the available evidence, contextualize evidence and recommendations, identify gaps, and suggest pragmatic solutions [77-79].

The second challenge is to select relevant and applicable evidence for people living in the community. In particular, this study stresses the challenge of selecting the evidence base for the prevention of a complication in the context of a complex chronic condition. Indeed, when selecting the prevention measures for PIs, experts have to take into consideration all aspects of self-management as well as feasibility issues. For instance, in the case of the prevention of PIs in individuals with SCI, it was mentioned during the working group that hydration is very important for preventing PIs; however, liquid intake often requires catheterization, which, in turn, can increase the risk of bladder infections. Similarly, doing pushup exercise to relieve the skin is good for preventing PIs, but it could cause damage to the shoulders in the long term. These examples illustrate the complexity and sometimes conflicting nature of evidence-based recommendations that are feasible for community-dwelling individuals and that ensure a comprehensive approach to the self-management of SCI. Indeed, systematic reviews and meta-analysis offer valuable synthesis of the evidence [80], but they often have a narrow focus (eg, one complication), and in many cases, they only report on experimental studies, which, owing to their rigor, avoid biases (eg, confounding factors and selection bias), but do not consider real-life situations [81]. In order to overcome these limitations and achieve a comprehensive approach to self-management, it is fundamental for experts from all relevant specialties as well as the persons affected by the health condition to be involved in the selection of the evidence for mHealth interventions. The combination of interdisciplinarity and lived experiences ensures that all perspectives are represented in the discussion. However, for the discussion to be constructive and achieve agreement on a shared decision, a structured process is needed. A consensus meeting represents a valid method to synthesize information and enable decisions to be made when published information is inconsistent or inadequate [82], and it is widely used in medical and health services research [83-85].

### Limitations

We have to acknowledge a few limitations of our study. The first one is related to the selection of recommendations, as we searched only for recommendations in English and German. We also focused on recommendations specific to SCI and PIs, not considering, for instance, other recommendations on SCI in general or on PIs in other populations. The second limitation is linked to the participants in the consensus meeting. All relevant stakeholders were represented; however, the participant mix could have been more balanced (eg, there were many nurses and only one occupational therapist). In addition, the consensus meeting was held on only one day. This resulted in focused discussions on many relevant aspects of the recommendations, but it was very intense for all participants. Having more time at our disposal could have also allowed an additional discussion and voting round to avoid concluding the meeting while still having some ambiguous recommendations, which the research team later needed to clarify. We also acknowledge that the procedure used has not been compared with another procedure and has not been evaluated. However, the commitment of the participants during the procedure showed that the participatory approach was positively received.

### Strengths

Although this study had the above-mentioned limitations, it is important to acknowledge some of its strengths. The methodological choice of holding a consensus meeting has been proven to be highly valuable, as its structured process guarantees a democratic discussion and a judicial model [53]; hence, it provides a viable and transparent option for a true participatory design process. Three other strengths that we want to highlight helped the procedures of stage 2 to run more smoothly. First, the selection of experts through other professionals and an informal network proved to be highly valuable, and it provided credibility to our invitation. Moreover, being aware of time constraints, we condensed the consensus meeting activities in one day and provided stakeholders with two dates as options. Second, it should be noted that for constructive discussions during the consensus meeting, participant preparation for the session was extremely important (ie, having read the preparatory document describing the procedure and the list of recommendations resulting from stage 1). Third, having an efficient and automated voting system was essential for ensuring that the results of one vote were quickly available for the next round (plenary or group discussion). This case study proved the value of the presented procedure; however, as this was a demonstration study, there is a requirement for further studies to validate the approach.

### Conclusion

Considering that people need evidence-based information to make informed decisions and participate in health [29], this study may be valuable for improving the quality of mHealth interventions as it detailed the participatory procedure needed for the selection of the scientific evidence that forms the basis of mHealth content. In particular, this procedure might be useful

in the selection of evidence-based content in the case of complex chronic health conditions, for which every recommendation needs to be evaluated and considered in light of all other self-management requirements. Hence, agreement among all experts and affected individuals on which evidence is to be included is essential.

## Acknowledgments

The authors would like to thank the Swiss Accident Insurance Fund (Suva) for funding the project this paper was based on. We are also very grateful to Nadia Lustenberger and the master's students who supported us in the literature review, Melissa Selb and Wolfgang Segerer for their advice and invaluable support in the organization and conduction of the consensus meeting, and other colleagues who assisted as note takers and facilitators. We would also like to thank all the participants for their valuable insights and collaboration.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
User personas.
[PDF File (Adobe PDF File), 616 KB - mhealth_v8i5e15818_app1.pdf ]

Multimedia Appendix 2
Final set of recommendations for pressure injury prevention and management.
[DOCX File , 24 KB - mhealth_v8i5e15818_app2.docx ]

## References

1. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
2. Olla P, Shimskey C. mHealth taxonomy: a literature survey of mobile health applications. Health Technol 2015 Jan 30;4(4):299-308. [doi: 10.1007/s12553-014-0093-8]
3. Fiordelli M, Diviani N, Schulz PJ. Mapping mHealth research: a decade of evolution. J Med Internet Res 2013;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]
4. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 01;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
5. Marcano Belisario JS, Huckvale K, Greenfield G, Car J, Gunn LH. Smartphone and tablet self management apps for asthma. Cochrane Database Syst Rev 2013 Nov 27(11):CD010013 [FREE Full text] [doi: 10.1002/14651858.CD010013.pub2] [Medline: 24282112]
6. Tung JY, Stead B, Mann W, Pham B, Popovic MR. Assistive technologies for self-managed pressure ulcer prevention in spinal cord injury: A scoping review. J Rehabil Res Dev 2015;52(2):131-146. [doi: 10.1682/jrrd.2014.02.0064]
7. Breland JY, Yeh VM, Yu J. Adherence to evidence-based guidelines among diabetes self-management apps. Transl Behav Med 2013 Sep 1;3(3):277-286 [FREE Full text] [doi: 10.1007/s13142-013-0205-4] [Medline: 24073179]
8. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
9. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
10. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert Involvement and Adherence to Medical Evidence in Medical Mobile Phone Apps: A Systematic Review. JMIR Mhealth Uhealth 2015 Jul 27;3(3):e79 [FREE Full text] [doi: 10.2196/mhealth.4169] [Medline: 26215371]
11. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
12. Aromataro O, Van Hoye A, Vuillemin A, Foucaut A, Crozet C, Pommier J, et al. How do mobile health applications support behaviour changes? A scoping review of mobile health applications relating to physical activity and eating behaviours. Public Health 2019 Oct;175:8-18 [FREE Full text] [doi: 10.1016/j.puhe.2019.06.011] [Medline: 31374453]
13. Hartin PJ, Nugent CD, McClean SI, Cleland I, Tschanz JT, Clark CJ, et al. The Empowering Role of Mobile Apps in Behavior Change Interventions: The Gray Matters Randomized Controlled Trial. JMIR Mhealth Uhealth 2016 Aug 02;4(3):e93 [FREE Full text] [doi: 10.2196/mhealth.4878] [Medline: 27485822]

14. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes Prevention and Weight Loss with a Fully Automated Behavioral Intervention by Email, Web, and Mobile Phone: A Randomized Controlled Trial Among Persons with Prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
15. Tombor I, Shahab L, Brown J, Crane D, Michie S, West R. Development of SmokeFree Baby: a smoking cessation smartphone app for pregnant smokers. Transl Behav Med 2016 Dec;6(4):533-545 [FREE Full text] [doi: 10.1007/s13142-016-0438-0] [Medline: 27699682]
16. Mummah SA, King AC, Gardner CD, Sutton S. Iterative development of Vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 08;13:90 [FREE Full text] [doi: 10.1186/s12966-016-0400-z] [Medline: 27501724]
17. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A Framework and Toolkit of Strategies for the Development of More Effective Digital Interventions to Change Health Behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
18. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
19. Cho Y, Lee S, Islam SM, Kim S. Theories Applied to m-Health Interventions for Behavior Change in Low- and Middle-Income Countries: A Systematic Review. Telemed J E Health 2018 Oct;24(10):727-741 [FREE Full text] [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
20. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. Cochrane Database Syst Rev 2017 Sep 25;9:CD011479 [FREE Full text] [doi: 10.1002/14651858.CD011479.pub2] [Medline: 28944453]
21. Modave F, Bian J, Leavitt T, Bromwell J, Harris Iii C, Vincent H. Low Quality of Free Coaching Apps With Respect to the American College of Sports Medicine Guidelines: A Review of Current Mobile Apps. JMIR Mhealth Uhealth 2015 Jul 24;3(3):e77 [FREE Full text] [doi: 10.2196/mhealth.4669] [Medline: 26209109]
22. Kebede M, Steenbock B, Helmer SM, Sill J, Möllers T, Pischke CR. Identifying Evidence-Informed Physical Activity Apps: Content Analysis. JMIR Mhealth Uhealth 2018 Dec 18;6(12):e10314 [FREE Full text] [doi: 10.2196/10314] [Medline: 30563810]
23. Knight E, Stuckey MI, Prapavessis H, Petrella RJ. Public health guidelines for physical activity: is there an app for that? A review of android and apple app stores. JMIR Mhealth Uhealth 2015 May 21;3(2):e43 [FREE Full text] [doi: 10.2196/mhealth.4003] [Medline: 25998158]
24. Breton ER, Fuemmeler BF, Abroms LC. Weight loss-there is an app for that! But does it adhere to evidence-informed practices? Transl Behav Med 2011 Dec;1(4):523-529 [FREE Full text] [doi: 10.1007/s13142-011-0076-5] [Medline: 24073074]
25. Lau Y, Cheng LJ, Chi C, Tsai C, Ong KW, Ho-Lim SS, et al. Development of a Healthy Lifestyle Mobile App for Overweight Pregnant Women: Qualitative Study. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e91 [FREE Full text] [doi: 10.2196/mhealth.9718] [Medline: 29685868]
26. White BK, Martin A, White JA, Burns SK, Maycock BR, Giglia RC, et al. Theory-Based Design and Development of a Socially Connected, Gamified Mobile App for Men About Breastfeeding (Milk Man). JMIR Mhealth Uhealth 2016 Jun 27;4(2):e81 [FREE Full text] [doi: 10.2196/mhealth.5652] [Medline: 27349756]
27. Wilhide Iii CC, Peeples MM, Anthony Kouyaté RC. Evidence-Based mHealth Chronic Disease Mobile App Intervention Design: Development of a Framework. JMIR Res Protoc 2016 Feb 16;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
28. Gellert P, Tille F. What do we know so far? The role of health knowledge within theories of health literacy. The European Health Psychologist 2015;17(6):266-274 [FREE Full text]
29. Edwards A, Elwyn G. Evidence-based Patient Choice. Oxford: Oxford University Press; 2001.
30. Haffey F, Brady RR, Maxwell S. A comparison of the reliability of smartphone apps for opioid conversion. Drug Saf 2013 Feb;36(2):111-117. [doi: 10.1007/s40264-013-0015-0] [Medline: 23322549]
31. Huckvale K, Car M, Morrison C, Car J. Apps for asthma self-management: a systematic assessment of content and tools. BMC Med 2012;10:144 [FREE Full text] [doi: 10.1186/1741-7015-10-144] [Medline: 23171675]
32. Kalz M, Lenssen N, Felzen M, Rossaint R, Tabuenca B, Specht M, et al. Smartphone apps for cardiopulmonary resuscitation training and real incident support: a mixed-methods evaluation study. J Med Internet Res 2014;16(3):e89 [FREE Full text] [doi: 10.2196/jmir.2951] [Medline: 24647361]
33. Biermann JS, Golladay GJ, Greenfield ML, Baker LH. Evaluation of cancer information on the Internet. Cancer 1999 Aug 01;86(3):381-390. [Medline: 10430244]
34. Smith H, Gooding S, Brown R, Frew A. Evaluation of readability and accuracy of information leaflets in general practice for patients with asthma. BMJ 1998 Jul 25;317(7153):264-265 [FREE Full text] [doi: 10.1136/bmj.317.7153.264] [Medline: 9677221]

35. Groussard P, Pigot H, Giroux S. From conception to evaluation of mobile services for people with head injury: A participatory design perspective. Neuropsychol Rehabil 2018 Jul;28(5):667-688. [doi: 10.1080/09602011.2015.1117499] [Medline: 26679473]
36. Noergaard B, Sandvei M, Rottmann N, Johannessen H, Wiil U, Schmidt T, et al. Development of a Web-Based Health Care Intervention for Patients With Heart Disease: Lessons Learned From a Participatory Design Study. JMIR Res Protoc 2017 May 17;6(5):e75 [FREE Full text] [doi: 10.2196/resprot.7084] [Medline: 28526674]
37. Hendriks N, Slegers K, Wilkinson A. Lessons Learned from Participatory Design in Dementia Care: Placing Care Partners at the Centre. Stud Health Technol Inform 2017;233:63-77. [Medline: 28125414]
38. Konnerup U. Engaging People with Aphasia in Design of Rehabilitation Through Participatory Design: A Way to Learn what They Really Want. Stud Health Technol Inform 2017;233:148-157. [Medline: 28125421]
39. Newman C, Shankar R, Hanna J, McLean B, Osland A, Milligan C, et al. Developing an Evidence-Based Epilepsy Risk Assessment eHealth Solution: From Concept to Market. JMIR Res Protoc 2016 Jun 07;5(2):e82 [FREE Full text] [doi: 10.2196/resprot.5389] [Medline: 27269382]
40. Baskerville NB, Struik LL, Dash D. Crush the Crave: Development and Formative Evaluation of a Smartphone App for Smoking Cessation. JMIR Mhealth Uhealth 2018 Mar 02;6(3):e52 [FREE Full text] [doi: 10.2196/mhealth.9011] [Medline: 29500157]
41. Birrell L, Deen H, Champion KE, Newton NC, Stapinski LA, Kay-Lambkin F, et al. A Mobile App to Provide Evidence-Based Information About Crystal Methamphetamine (Ice) to the Community (Cracks in the Ice): Co-Design and Beta Testing. JMIR Mhealth Uhealth 2018 Dec 20;6(12):e11107 [FREE Full text] [doi: 10.2196/11107] [Medline: 30573443]
42. International Spinal Cord Society. World Health Organization. Geneva: World Health Organization; 2013. International perspectives on spinal cord injury URL: https://www.who.int/disabilities/policies/spinal_cord_injury/en/ [accessed 2020-05-12]
43. Sweis R, Biller J. Systemic Complications of Spinal Cord Injury. Curr Neurol Neurosci Rep 2017 Feb;17(2):8. [doi: 10.1007/s11910-017-0715-4] [Medline: 28188542]
44. Brinkhof M, Al-Khodairy A, Eriks-Hoogland I, Fekete C, Hinrichs T, Hund-Georgiadis M, SwiSCI Study Group. Health conditions in people with spinal cord injury: Contemporary evidence from a population-based community survey in Switzerland. J Rehabil Med 2016 Feb;48(2):197-209 [FREE Full text] [doi: 10.2340/16501977-2039] [Medline: 26931074]
45. Regan MA, Teasell RW, Wolfe DL, Keast D, Mortenson WB, Aubut JL, Spinal Cord Injury Rehabilitation Evidence Research Team. A systematic review of therapeutic interventions for pressure ulcers after spinal cord injury. Arch Phys Med Rehabil 2009 Feb;90(2):213-231 [FREE Full text] [doi: 10.1016/j.apmr.2008.08.212] [Medline: 19236976]
46. Byrne DW, Salzberg CA. Major risk factors for pressure ulcers in the spinal cord disabled: a literature review. Spinal Cord 1996 May;34(5):255-263. [doi: 10.1038/sc.1996.46] [Medline: 8963971]
47. Consortium for Spinal Cord Medicine Clinical Practice Guidelines. Pressure ulcer prevention and treatment following spinal cord injury: a clinical practice guideline for health-care professionals. J Spinal Cord Med 2001;24 Suppl 1:S40-101. [doi: 10.1080/10790268.2001.11753592] [Medline: 11958176]
48. Bogie KM, Reger SI, Levine SP, Sahgal V. Electrical stimulation for pressure sore prevention and wound healing. Assist Technol 2000;12(1):50-66. [doi: 10.1080/10400435.2000.10132009] [Medline: 11067577]
49. Houlihan BV, Jette A, Paasche-Orlow M, Wierbicky J, Ducharme S, Zazula J, et al. A telerehabilitation intervention for persons with spinal cord dysfunction. Am J Phys Med Rehabil 2011 Sep;90(9):756-764. [doi: 10.1097/PHM.0b013e31820b140f] [Medline: 21389846]
50. Kruger EA, Pires M, Ngann Y, Sterling M, Rubayi S. Comprehensive management of pressure ulcers in spinal cord injury: current concepts and future trends. J Spinal Cord Med 2013 Nov;36(6):572-585 [FREE Full text] [doi: 10.1179/2045772313Y.0000000093] [Medline: 24090179]
51. Schubart JR, Hilgart M, Lyder C. Pressure ulcer prevention and management in spinal cord-injured adults: analysis of educational needs. Adv Skin Wound Care 2008 Jul;21(7):322-329. [doi: 10.1097/01.ASW.0000323521.93058.47] [Medline: 18600074]
52. Cogan AM, Blanchard J, Garber SL, Vigen CL, Carlson M, Clark FA. Systematic review of behavioral and educational interventions to prevent pressure ulcers in adults with spinal cord injury. Clin Rehabil 2017 Jul;31(7):871-880. [doi: 10.1177/0269215516660855] [Medline: 27440806]
53. Bourrée F, Michel P, Salmi L. Consensus methods: Review of original methods and their main alternatives used in public health. Revue d'Épidémiologie et de Santé Publique 2008 Dec;56(6):e13-e21. [doi: 10.1016/j.respe.2008.10.005]
54. Holden MA, Haywood KL, Potia TA, Gee M, McLean S. Recommendations for exercise adherence measures in musculoskeletal settings: a systematic review and consensus meeting (protocol). Syst Rev 2014 Feb 10;3(1):10 [FREE Full text] [doi: 10.1186/2046-4053-3-10] [Medline: 24512976]
55. Stucki A, Cieza A, Michel F, Stucki G, Bentley A, Culebras A, et al. Developing ICF Core Sets for persons with sleep disorders based on the International Classification of Functioning, Disability and Health. Sleep Med 2008 Jan;9(2):191-198. [doi: 10.1016/j.sleep.2007.01.019] [Medline: 17644416]

56. Woods L, Cummings E, Duff J, Walker K, editors. The development use of personas in a user-centred mHealth design project. In: Proceedings of the 29th Australian Conference on Computer-Human Interaction. 2017 Nov Presented at: Australian Conference on Computer-Human Interaction: ACM; 2017; Australia p. 560-565. [doi: 10.1145/3152771.3156186]
57. Barberà-Guillem R, Campos N, Biel S, Erdt S, Payá J, Ganzarain J. User involvement: how we integrated users into the innovation process and what we learned from it. In: Assistive Technologies for the Interaction of the Elderly. Berlin, Germany: Springer; 2014:33-47.
58. Böthig R, Domurath B, Kaufmann A, Bremer J, Vance W, Kurze I. [Neuro-urological diagnosis and therapy of lower urinary tract dysfunction in patients with spinal cord injury : S2k Guideline of the German-Speaking Medical Society of Paraplegia (DMGP), AWMF register no. 179/001]. Urologe A 2017 Jun 17;56(6):785-792. [doi: 10.1007/s00120-017-0354-z] [Medline: 28314966]
59. Eisenhuth J, Geyh S, Gottschalk S, Kues S, Neikes M, Nüsslein ST. Psychologische Aspekte in der Dekubitusprophylaxe. Empfehlungen des Arbeitskreises Psychologie der Deutschsprachigen Medizinischen Gesellschaft für Paraplegie (DMGP) 2012.
60. Houghton P, Campbell K. Canadian best practice guidelines for the prevention and management of pressure ulcers in people with Spinal Cord Injury: a resource handbook for clinicians. Ontario: Ontario Neurotrauma Foundation; 2013:a.
61. Bowman T. Ontario Neurotrauma Foundation. Preventing and treating pressure sores: A guide for people with spinal cord injury URL: https://onf.org/wp-content/uploads/2019/04/Pressure_Ulcer_Guide_medium-res_single_pages.pdf [accessed 2017-10-20]
62. National Pressure Ulcer Advisory Panel. Prevention and treatment of pressure ulcers: quick reference guide. Cambridge: Cambridge Media; 2014.
63. Hsieh J, McIntyre A, Wolfe D, Lala D, Titus L, Campbell K. Pressure ulcers following spinal cord injury. Spinal Cord Injury Rehabilitation Evidence Version 2014;5:1-90.
64. Jain A. ISCOS - Textbook on comprehensive management of spinal cord injuries. Indian J Orthop 2016;50(2):223. [doi: 10.4103/0019-5413.177576]
65. Schweizer Paraplegiker Zentrum. Patient education, Druckstellen-Dekubitus v1. Schweizer Peraplegiker Zentrum education 2017. [doi: 10.1163/2352-0248_edn_a4880000]
66. Keast DH, Parslow N, Houghton PE, Norton L, Fraser C. Best Practice Recommendations for the Prevention and Treatment of Pressure Ulcers. Advances in Skin & Wound Care 2007;20(8):447-460. [doi: 10.1097/01.asw.0000284922.69932.c5]
67. Glaser BG. The Constant Comparative Method of Qualitative Analysis. Social Problems 1965 Apr;12(4):436-445. [doi: 10.1525/sp.1965.12.4.03a00070]
68. Laughton GE, Buchholz AC, Martin Ginis KA, Goy RE, SHAPE SCI Research Group. Lowering body mass index cutoffs better identifies obese persons with spinal cord injury. Spinal Cord 2009 Oct;47(10):757-762. [doi: 10.1038/sc.2009.33] [Medline: 19350042]
69. Sopher R, Nixon J, Gorecki C, Gefen A. Exposure to internal muscle tissue loads under the ischial tuberosities during sitting is elevated at abnormally high or low body mass indices. J Biomech 2010 Jan 19;43(2):280-286. [doi: 10.1016/j.jbiomech.2009.08.021] [Medline: 19762029]
70. Spreyermann R. Vor-und Nachsorge II: Gesundheitscoaching bei querschnittgelähmten Patientinnen und Patienten. 2013. URL: https://ssop.ch/assets/Uploads/Vorsorge-u-Nachsorge-II.pdf [accessed 2017-10-20]
71. Weaver FM, Collins EG, Kurichi J, Miskevics S, Smith B, Rajan S, et al. Prevalence of obesity and high blood pressure in veterans with spinal cord injuries and disorders: a retrospective review. Am J Phys Med Rehabil 2007 Jan;86(1):22-29. [doi: 10.1097/phm.0b013e31802b8937] [Medline: 17304685]
72. Wong S, Derry F, Jamous A, Hirani SP, Grimble G, Forbes A. Validation of the spinal nutrition screening tool (SNST) in patients with spinal cord injuries (SCI): result from a multicentre study. Eur J Clin Nutr 2012 Mar 14;66(3):382-387. [doi: 10.1038/ejcn.2011.209] [Medline: 22166898]
73. American Diabetes Association. Academy of Nutrition and Dietetics. Evidence-Based Nutrition Practice Guideline URL: https://www.andeal.org/category.cfm?cid=14 [accessed 2018-02-11]
74. Schweizerische Gesellschaft für Ernährung. 2019. URL: http://www.sge-ssn.ch/ [accessed 2018-02-11]
75. Groah SL, Schladen M, Pineda CG, Hsieh CJ. Prevention of Pressure Ulcers Among People With Spinal Cord Injury: A Systematic Review. PM R 2015 Jun 18;7(6):613-636. [doi: 10.1016/j.pmrj.2014.11.014] [Medline: 25529614]
76. Atkinson RA, Cullum NA. Interventions for pressure ulcers: a summary of evidence for prevention and treatment. Spinal Cord 2018 Mar 25;56(3):186-198. [doi: 10.1038/s41393-017-0054-y] [Medline: 29371701]
77. Hofmeijer J. Evidence-based medical knowledge: the neglected role of expert opinion. J Eval Clin Pract 2014 Dec;20(6):803-808. [doi: 10.1111/jep.12267] [Medline: 25267262]
78. Ehrich J, Somekh E, Pettoello-Mantovani M. The Importance of Expert Opinion-Based Data: Lessons from the European Paediatric Association/Union of National European Paediatric Societies and Associations (EPA/UNEPSA) Research on European Child Healthcare Services. J Pediatr 2018 Apr;195:310-311.e1. [doi: 10.1016/j.jpeds.2017.12.064] [Medline: 29643017]

79. Ponce OJ, Alvarez-Villalobos N, Shah R, Mohammed K, Morgan RL, Sultan S, et al. What does expert opinion in guidelines mean? A meta-epidemiological study. Evid Based Med 2017 Oct 18;22(5):164-169. [doi: 10.1136/ebmed-2017-110798] [Medline: 28924055]
80. Bartolucci AA, Hillegass WB. Overview, Strengths, and Limitations of Systematic Reviews and Meta-Analyses. In: Chiappelli F, editor. Evidence-Based Practice: Toward Optimizing Clinical Outcomes. Berlin, Heidelberg: Springer; 2010:17-33.
81. Wallace J, Nwosu B, Clarke M. Barriers to the uptake of evidence from systematic reviews and meta-analyses: a systematic review of decision makers' perceptions. BMJ Open 2012 Sep 01;2(5):e001220 [FREE Full text] [doi: 10.1136/bmjopen-2012-001220] [Medline: 22942232]
82. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ 1995 Aug 05;311(7001):376-380 [FREE Full text] [doi: 10.1136/bmj.311.7001.376] [Medline: 7640549]
83. European Society of Paedriatric Endocrinology. 2017. Consensus Meeting and Guideline Development URL: https://www.eurospe.org/clinical-practice/consensus-meeting-and-guideline-development/ [accessed 2019-11-12]
84. World Health Organization. 2017. Consensus meeting report: development of a target product profile ( TPP) and a framework for evaluation for a test for predicting progression from tuberculosis infection to active disease URL: https://apps.who.int/iris/handle/10665/259176 [accessed 2020-04-19]
85. Morris C, Janssens A, Allard A, Thompson CJ, Shilling V, Tomlinson R. Informing the NHS Outcomes Framework: evaluating meaningful health outcomes for children with neurodisability using multiple methods including systematic review, qualitative research, Delphi survey and consensus meeting. In: Health Services and Delivery Research. Southampton: NIHR Journals Library; 2014.


## Abbreviations

**eHealth:** electronic health
**mHealth:** mobile health
**PI:** pressure injury
**SCI:** spinal cord injury


*Edited by G Eysenbach; submitted 08.08.19; peer-reviewed by F Beyer, A Berrocal, S Lee, H Kondylakis; comments to author 21.01.20; revised version received 17.03.20; accepted 25.03.20; published 20.05.20.*

*Please cite as:*
*Fiordelli M, Zanini C, Amann J, Scheel-Sailer A, Brach M, Stucki G, Rubinelli S*
*Selecting Evidence-Based Content for Inclusion in Self-Management Apps for Pressure Injuries in Individuals With Spinal Cord Injury: Participatory Design Study*
*JMIR Mhealth Uhealth 2020;8(5):e15818*
*URL: http://mhealth.jmir.org/2020/5/e15818/*
*doi:10.2196/15818*
*PMID:32432559*

Original Paper

# Peer-to-Peer Social Media as an Effective Prevention Strategy: Quasi-Experimental Evaluation

William Evans[1], PhD; Elizabeth Andrade[1], DrPH; Michaela Pratt[2], BA; Alexandra Mottern[2], MS; Sergio Chavez[2], BA; Anthony Calzetta-Raymond[1], BA; Jiayan Gu[1], MS

[1]Milken Institute School of Public Health, George Washington University, Washington, DC, United States

[2]Mentor Foundation USA, Tysons, VA, United States

**Corresponding Author:**
William Evans, PhD
Milken Institute School of Public Health
George Washington University
Washington, DC
United States
Phone: 1 2023519546
Email: wdevans@gwu.edu

## Abstract

**Background:** Substance use by adolescents remains to be at unacceptably high levels, and there is evidence that teens' social norms are becoming more favorable toward recreational use and perceived safety of substances such as marijuana and prescription opioids. Social media offer a low-cost, potentially high-impact approach to disseminate prevention messages.

**Objective:** Living the Example (LTE) is a program that trains adolescent youth ambassadors to develop and disseminate prevention messages within their own social media networks and through in-school activities. This study aimed to evaluate the effects of exposure to LTE-based social media on students in the youth ambassadors' networks.

**Methods:** The George Washington (GW) University designed and implemented a quasi-experimental evaluation of the LTE program in 3 Maryland high schools. Before program launch, a sample of 826 students (wave 1) at the 3 schools, drawn from a census of freshmen enrolled in a class attended by all students at the grade level, completed a survey. A total of 584 students were surveyed at the wave 2 program midpoint and 542 at the wave 3 endpoint. The survey contained questions on drug use–related attitudes, beliefs, intentions, and behaviors, all based on validated measures. We evaluated the effects of LTE on the intended next 30-day drug use, and controlling for LTE self-reported exposure, age, and gender from waves 2 and 3 was appended into a single dataset. We first conducted ordinal logistic regressions for each drug use intention in wave 3 (ie, sell or distribute illegal drugs, smoke cigarettes, drink beer/wine/hard liquor when parents do not know about it, use marijuana, use lysergic acid diethylamide, cocaine, amphetamines or other illegal drugs, use heroin, use synthetic drugs, and use any prescription pills without a prescription) to examine the association between LTE exposure and drug use intentions. We included an interaction term for the study wave to examine intervention effects.

**Results:** We found a significant positive effect of LTE exposure on all 8 measured drug use intentions: sell/distribute illegal drugs; smoke cigarettes; drink beer, wine, or liquor when my parents do not know about it; use marijuana; use cocaine, amphetamines, or other illegal drug; use heroin; use synthetic drugs; use any prescription pills without a prescription (all $P<.05$; odds ratios ranging from 2.12 to 3.71). We also found that boys were more likely than girls to exhibit reduced drug use intentions. We also found reductions in 30-day intentions between the second and third survey waves for all 8 measured drug use variables.

**Conclusions:** Overall, the results are consistent with and indicate a stronger LTE effect in this study compared with a previous pilot study. LTE appears to offer a protective effect, with exposure to program messages leading to reduced/improved drug use intentions.

***(JMIR Mhealth Uhealth 2020;8(5):e16207)*** doi:10.2196/16207

**KEYWORDS**

social media; substance use; prevention; marijuana; opioids; adolescent health

# Introduction

## Background

As shown in the most recent Monitoring the Future study, substance abuse is among the greatest public health threats [1]. Substance use by adolescents remains to be at unacceptably high levels, and there is evidence that teens' social norms are becoming more favorable toward recreational use and perceived relative safety of substances such as marijuana and prescription opioids [2-4]. Teens are particularly sensitive to the messaging about drugs in their social environment, including in digital communities, and drug use choices are driven by availability, acceptability, and perceived risk [5,6]. Recently there have been many changes in the availability of marijuana to the public. The consequences of use and the situations in which individuals may choose to use marijuana are changing in many states owing to medical marijuana and legalization [7], potentially influencing youth perceptions of acceptability and risk [8]. Furthermore, due to widespread availability and a landscape of more dangerous drugs, including fentanyl contamination, opioid overdose death rates continue to rise despite overall declines in use [9,10]. Yet, some US $40 billion per year is spent on prevention programs [11], including major efforts involving health communication, such as the Above the Influence campaign [12]. Given this growing public health threat, prevention efforts must prioritize engagement strategies that address the evolving substance use risk perceptions and leverage youths' and young adults' affinity for digital technology [13].

There is growing evidence that marijuana use has negative health consequences for adolescents, especially when use begins early and when combined with other substance use [14]. Recent studies suggest adolescent marijuana use may be linked to altered longer-term neurodevelopmental trajectories and compromised neural health, impaired frontal lobe function, and psychosocial effects [14-16]. Additionally, early-onset adolescent marijuana use combined with alcohol and other substance use has been linked to numerous cognitive impairments and neural health effects [17,18]. Social norms favoring marijuana use are increasing [19,20] and may be associated with the relaxation of marijuana laws [21,22]. At the same time, both natural and synthetic opioid use has clear and well-documented health consequences, and adolescents are at a heightened risk for long-term neurodevelopmental effects [23]. Approximately 68% of the 70,237 drug overdose deaths in 2017 involved opioids, and over 28,466 of those deaths were from synthetic opioids such as fentanyl [24].

## Social Media and Prevention

Social media and social networks are major drivers of public debate and perceptions of marijuana, opioids, and other drugs [7,14,15,25,26], including for the youth, and there are numerous messages promoting the acceptability of substance use in these digital networks. This highlights the need for novel interventions that use digital prevention strategies and offer a counternarrative or counterargument [27]. Recent studies have examined exposure to antidrug communications and found that counternarrative strategies may partially explain positive campaign outcomes [28,29]. A campaign's *brand marketing strategy* may in part determine how the audience responds to antidrug advertising [12]. Substance abuse prevention programs and behaviors promoted by these programs can have brand identities as well. Recent studies have linked drug resistance branding to improved substance use outcomes [30,31]. Our research team has pioneered branding for prevention in multiple domains [32,33]. In the context of relaxed marijuana laws and heightened opioid risk, this study investigated how prevention can be optimized through branded peer-to-peer messaging and digital engagement.

The Mentor Foundation USA implements a program entitled, Shatter the Myths (STM) Youth Rallies—this program conducts interactive in-person events with high school students and is designed to raise an awareness of risks and dispel myths surrounding drug use [34-37]. STM Youth Rallies are based on recent neuroscience, use an experiential learning approach to prevention, and apply principles from branding and social marketing [38,39]. By focusing on youths' innate talents and strengths, the rallies enable youths to become advocates regarding the benefits of living a drug-free lifestyle.

Mentor Foundation USA is part of the Mentor International organization, which implements youth development programs worldwide. Mentor International's work is guided by the Positive Youth Development (PYD) theory [40], working with children and families to combat risk behaviors and promote healthy lifestyles. Mentor International follows the United Nations Sustainable Development Goals, in particular, by ensuring healthy lives and promoting well-being for all ages.

Living the Example (LTE) has been designed based on the PYD theory, following the authors' previous research, to address the rapidly changing substance use social environment in the United States by empowering the youth to serve as peer change agents [41,42]. LTE builds upon the STM Youth Rallies by adding a social media–delivered component to spread peer-to-peer substance use prevention messages, while also applying social marketing and branding approaches.

LTE represents a novel approach to addressing a wide range of substances used by adolescents through gain-framed messaging and digital engagement. According to the Prospect Theory, framing warnings to emphasize the negative health effects of marijuana and opioid use (ie, loss-framed) or benefits of avoiding marijuana and opioid use (ie, gain-framed) will differentially affect prevention outcomes [43,44]. Research on message framing [45-47] and prevention messages for preteens and teenagers [48-50] suggests that gain-framed messages may be effective. Specifically, research on smoking cessation messaging favors gain-framed messages about prevention and health-promoting behaviors [51-53]. For example, the communication theory indicates that gain-framed messages conveying the health benefits of quitting smoking may be optimal to promote cessation [54]. Such messages have the potential to produce behavior change through mediated pathways, such as enhancing beliefs that quitting reduces risks and reducing the attractiveness of industry branding [55,56]. Similar gain-framed messages can be adapted to marijuana and opioid use prevention to address the changing prevention environment. The use of graphic imagery has also been found

to contribute to the uptake of gain-framed health messages [57,58]. Social media are effective tools to share graphic, multimedia prevention content that may enhance framing effects. LTE is promising in that it applies this approach to prevention by promoting the use of positive antidrug brand representations delivered via peer networks on social media.

### Living the Example Intervention Research

In this study, we adapted and evaluated a novel intervention, LTE, which has been previously pilot tested and demonstrated to be effective in reducing 30-day marijuana and prescription painkiller use intentions [42]. The LTE pilot was based on an after-school curriculum for *youth ambassadors*, followed by *ambassador* promotion of peer substance use avoidance with user-generated content via digital social networks and through school-wide *change projects*. LTE addresses numerous forms of substance use overall, including all widely used illicit drugs and alcohol, and the evaluation includes measures of intentions and use. This study focused specifically on marijuana and opioid use due to the rapid growth and prevalence of these substances and because the previous pilot demonstrated effectiveness in reducing use intentions.

The specific aim of this research was to implement and evaluate the adapted LTE curriculum in 3 suburban Maryland high schools. We tested 3 hypotheses:

1. LTE *youth ambassador* training would be successful in disseminating prevention messages via social media during the school year.
2. Exposure to LTE would be associated with lower 30-day substance use intentions at follow-up.
3. Higher levels of exposure to LTE would be associated with a reduction in 30-day substance use intentions across the program implementation period.

## Methods

### Design

We employed a quasi-experimental design (QED) through 3 cross-sectional surveys of freshmen at each school to test dose-response effects of exposure to LTE social media and school-based activities on substance use outcomes. We collected self-reported data on outcomes through questionnaires administered at 3 waves before, during, and at the conclusion of the program in the 3 high schools over a period of 9 months. We also collected data from the social media platforms used by students, such as Instagram, to independently measure their LTE posting activity.

### Intervention

The LTE curriculum was delivered as a series of structured after-school group sessions of 90-min each, 1 session per week. The sessions were delivered by a single program staff member following this outline:

- Session 1: What is a brand? Describe the idea of branding and branding substance use prevention.
- Session 2: Introduction to social media. Learn basics of social media, how it can influence message recipients, and how to create influential messages.
- Session 3: Boosting online engagement. How to connect and build engagement with social networks.
- Session 4: Using your voice—introduction to advocacy. Learn how to share opinions about an issue in the community with aim of influencing peers.
- Session 5: Advocacy in action. Develop skills to advocate for substance use prevention with peers.
- Session 6: Applying what we have learned. *Ambassadors* develop live content, share it with their social media networks, and engage with their peers.

Program staff, who were counselors and other school officials, recruited *youth ambassadors* at the 3 schools (46 in total), all seniors, at the start of the school year. The LTE curriculum sessions were then delivered through in-school sessions with *ambassadors* at school-assigned times. Each session was followed by staff-led questions and discussion topics to ensure *ambassador* comprehension of the content.

Following the last session, *ambassadors* were given 2 weeks to develop prevention content and share it live with their social media networks. This represented the launch of LTE prevention messaging to the broader student population beyond the *ambassador* group. After the initial *ambassador* training and launch of peer-to-peer prevention messaging, the program staff maintained weekly in-person, SMS text message, email, and instant messaging contact with *ambassadors* to encourage regular social media posting and provide help and encouragement in a coaching function. Additional activities, projects, and competitions were scheduled to keep *youth ambassadors* engaged in the LTE program throughout the course of the academic year. Competitions offered opportunities for youths to compete against each other and win prizes for their social media content and peer outreach. This strategy rewarded *ambassadors* for creative outreach techniques, effective inclusion of prevention content, persuasive and engaging tactics, compelling narrative stories, and other noteworthy features of their work. These competitions were also intended to boost intervention reach and amplify prevention messaging—for example, prizes were awarded for generating the most social media engagement (ie, shares, reactions, or comments) and using the *#livingtheexample* hashtag.

After completing the initial 6-week training curriculum, each participating school submitted a project description and budget to receive a monetary stipend to complete a “change project.” *Youth ambassadors* were challenged to design and implement change projects to increase peer knowledge and awareness about the harmful effects of substance use and address social norms, stigma, and perceptions about substance use and addiction. Approved and funded projects were implemented during the spring semester. Following successful project completion, each school submitted results of their work in the form of photos/videos and reports. These change projects were part of the overall LTE intervention and were captured in social media posts by *youth ambassadors;* measures of LTE exposure are inclusive of this activity.

In the spring of 2019, the program staff led interactive STM substance use prevention youth rallies at each participating school. Large numbers of students attended each rally, which

were held as school-wide assemblies. We do not have data on exact attendance. The purpose was to draw attention to the overall substance use prevention program, generate enthusiasm surrounding the *youth ambassadors*, and stimulate interest in *ambassador* LTE social media posts. We note that this was a 1-time event and served to reinforce the ongoing program. The STM rally was designed to dispel marijuana and opioid use myths (ie, that most peers use marijuana or that opioid use is safe) and to encourage positive peer influence. The STM rally focused on youths' talents and strengths, enabling them to become prevention advocates. The STM rally featured thought-provoking speakers, including *youth ambassadors*, a scientist from the National Institute on Drug Abuse (NIDA) who educated students about the science behind substance use and the adolescent brain, and a young person in recovery sharing an honest testimonial.

### Measures and Instruments

We adapted a previously tested questionnaire using validated scales from previous work by the authors [59,60], as well as from other validated scales from both the Substance Abuse and Mental Health Services Administration (SAMHSA) 2014 Communities that Care survey instrument and the 2012 Monitoring the Future survey [61]. The 88-item instrument was programmed into the SurveyMonkey software for computer-administered completion during a required freshmen English class at each of the high schools and took an average of 15 min to complete. In addition to demographic information and last grade completed in school, other scales used included the following: traditional and digital media use; attitudes toward social media; drug use risk perceptions; personal and perceived peer reasons to use drugs; drug use social norms; perceived peer drug use; reported peer drug use; self-reported past 30-day drug use; next 30-day drug use intentions; drug use/refusal influences; and self-reported exposure to the LTE program (ie, self-reported frequency of receiving social media messaging and from whom) and receptivity to the LTE program messages (ie, self-reported trust and how much they liked the program messages). The response option format for the drug use intention measure was on a 5-point Likert scale (ranging from strongly agree to strongly disagree) with statements about drug use intentions.

Additionally, we documented posts from a total of 24 (of the 46 total) *youth ambassadors* across the 3 high schools to social media platforms of their choice during an 8-month period, primarily on Instagram and Snapchat. We measured *ambassador* social media outreach activity, *ambassador* followers, type of outreach (ie, post, story, snap), and peer engagement (ie, likes, comments).

### Data Collection

Working with a local liaison at each high school, we scheduled the questionnaires to be first administered before the initiation of LTE *youth ambassador* training in November 2018 (baseline or wave 1). The required freshmen classes were invited to participate in the survey, which included a total potential sample of 923 freshmen students across the 3 high schools. We surveyed a total of 826 students at baseline (89.5% of the student census) across the 3 high schools during October and early November 2018. We did not record the students' personally identifiable information, as the study design was cross-sectional. The training was then implemented in 6 weekly sessions followed by an *ambassador* contest to launch LTE. We completed a midline (wave 2) survey in February to March 2019 with a total of 584 students recruited from the same classes as baseline. Finally, we completed the endline (wave 3) survey in June 2019 with a total of 542 students in the same classes. Thus, at wave 3, we were able to survey 58.7% (542/923) of all freshmen and 65.6% (542/826) of those surveyed at wave 1. Ambassadors were not sampled as part of survey data collection. We cannot independently confirm whether lower rates of completed surveys at follow-up were due to student absences on the survey date or refusal to participate.

We also collected social media data from *youth ambassador* posts using Hootsuite software and captured *ambassador* posts with the included *#livingtheexample* hashtag. Owing to the nature of the Instagram platform, stories are only available for a 24-hour period; in some cases, program staff requested that *ambassadors* provide screenshots of their posts for documentation purposes.

### Data Analysis

Drug use intentions, LTE exposure, and demographic (age and gender) data from waves 2 and 3 were appended into a single dataset. We first estimated ordinal logistic regression models for each drug use intention variable in wave 3 (ie, sell/distribute illegal drugs, smoke cigarettes, drink beer/wine/hard liquor when parents do not know about it, use marijuana, use lysergic acid diethylamide (LSD)/cocaine/amphetamines or other illegal drugs, use heroin, use synthetic drugs, use any prescription pills without a prescription) to examine the association between LTE exposure and drug use intentions. We assumed that the relationship between each pair of outcome categories was the same, meaning that the coefficients that describe the relationship between strongly agree versus all higher categories of the response are the same as those that agree versus all higher categories and so on.

In a separate set of regressions, we included an interaction term for the study wave to examine whether participants had stronger disagreement with drug use after the intervention was implemented. All regression models controlled for participant's age and gender. All data analysis was conducted in R version 3.5.3 (R Core Team, 2019). All regressions fit model assumptions have been described.

## Results

We recorded LTE social media activity from a total of 24 (of the 46) *youth ambassadors*. The Instagram story function was the most widely used by *ambassadors*, with a total of 45 stories, followed by 28 Instagram posts made during the program. The *ambassadors* had a total of 12,894 followers in their social media networks during the program period. There were 1291 follower likes for posts by *ambassadors* and 33 comments made by followers, likely a partial representation of engagement. It is worth noting that we cannot confirm that we captured all social media posts created by the *ambassadors* as some students may not have used the hashtag or provided complete data to the

coordinator on their posting activity. It should also be noted that, due to the nature of social media platforms, we cannot confirm the identity of individuals in the *youth ambassadors'* networks who engaged with the posts or how many of them represent unique individuals.

In the outcome surveys, we collected limited demographic information, and 52% of respondents were female, 55% were aged 14 years and 45% were aged 15 years, 96% were born in the United States, and all were freshmen. The survey focused mainly on substance use attitudes, beliefs, intentions, and behavior. Mean responses to the drug use intention scales for the outcomes of interest are shown in Table 1.

**Table 1.** Overall mean differences in drug use intentions among all baseline, midline, and endline groups.

| Measures | Baseline (n=835), mean (SD) | Midline (n=593), mean (SD) | Endline (n=557), mean (SD) | *P* value |
|---|---|---|---|---|
| Sell/distribute illegal drugs[a] | 3.69 (0.68) | 3.58 (0.82) | 3.54 (0.83) | <.001 |
| Smoke cigarettes | 3.72 (0.60) | 3.67 (0.69) | 3.57 (0.81) | <.001 |
| Drink beer, wine, or hard liquor (eg, vodka, whiskey, or gin) when my parents do not know about it | 3.56 (0.77) | 3.47 (0.85) | 3.36 (0.93) | <.001 |
| Use marijuana | 3.56 (0.81) | 3.48 (0.88) | 3.36 (0.99) | <.001 |
| Use LSD, cocaine, amphetamines, or other illegal drugs | 3.74 (0.62) | 3.72 (0.65) | 3.63 (0.77) | .008 |
| Use heroin | 3.75 (0.61) | 3.72 (0.66) | 3.65 (0.77) | .013 |
| Use synthetic drugs (eg, K2, spice, bath salts) | 3.75 (0.60) | 3.74 (0.64) | 3.63 (0.77) | .002 |
| Use prescription pills without a prescription | 3.71 (0.65) | 3.66 (0.73) | 3.59 (0.80) | .007 |

[a]For each of the following questions, please indicate whether you strongly agree=1, agree=2, neither agree nor disagree=3, disagree=4, or strongly disagree=5.

Next, we estimated ordinal logistic regression models for each of the 30-day substance use intention outcomes of interest (Table 2). In this model, we found a significant positive effect of LTE exposure on all 8 measured drug use intention variables: sell/distribute illegal drugs; smoke cigarettes; drink beer, wine or liquor when my parents do not know about it; use marijuana; use LSD, cocaine, amphetamines, or other illegal drug; use heroin; use synthetic drugs; use any prescription pills without a prescription (all $P<.05$; odds ratios ranging from 2.12 to 3.71). We also found that boys were more likely than girls to exhibit reductions in drug use intentions.

**Table 2.** Logistic regressions of drug use intentions on Living the Example program exposure at wave 3, exponentiated coefficient (n=542).

| Measures | Self-reported drug use intention, exponentiated coefficient (*P* value) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1[a] | 2[b] | 3[c] | 4[d] | 5[e] | 6[f] | 7[g] | 8[h] |
| LTE[i] exposure | 2.96 (<.003) | 3.42 (<.001) | 2.12 (<.007) | 2.91 (<.003) | 3.61 (<.001) | 3.58 (<.001) | 3.71 (<.001) | 3.44 (<.001) |
| Age (years) | 0.99 (.94) | 0.85 (.33) | 1.08 (.56) | 1.23 (.15) | 0.94 (.73) | 0.81 (.23) | 0.93 (.64) | 0.98 (.90) |
| Gender (female; reference) | 2.72 (<.006) | 2.71 (<.007) | 1.94 (<.009) | 2.74 (<.006) | 3.60 (<.001) | 3.24 (<.003) | 3.03 (<.004) | 2.76 (<.006) |

[a]Sell/distribute illegal drugs.

[b]Smoke cigarettes.

[c]Drink beer, wine, or hard liquor (eg, vodka, whiskey, or gin) when my parents do not know about it.

[d]Use marijuana.

[e]Use LSD, cocaine, amphetamines, or other illegal drugs.

[f]Use heroin.

[g]Use synthetic drugs (eg, K2, spice, bath salts).

[h]Use any prescription pills without a prescription.

[i]LTE: Living the Example.

Next, we estimated a separate model that included a survey wave interaction term to examine the hypothesis that substance use intentions declined during the LTE program as a function of exposure to peer-to-peer prevention messages via social media. As hypothesized, we observed reductions in 30-day intentions to use all 8 measured drugs between the second and third follow-up surveys, the period during which the social media messaging component of the LTE program took place (Table 3).

**Table 3.** Logistic regressions of drug use intentions on Living the Example program exposure and survey waves 2 to 3, exponentiated coefficient.

| Measures | Self-reported drug use intention, exponentiated coefficient (*P* value) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 1[a] | 2[b] | 3[c] | 4[d] | 5[e] | 6[f] | 7[g] | 8[h] |
| LTE[i] exposure | 3.31 (<.001) | 3.30 (<.001) | 2.43 (<.008) | 2.58 (<.007) | 3.33 (<.001) | 3.73 (<.001) | 3.57 (<.001) | 3.45 (<.001) |
| Wave 2 (reference) | 1.06 (.66) | 1.29 (.07) | 1.16 (.24) | 1.14 (.32) | 1.30 (.09) | 1.21 (.21) | 1.41 (.02) | 1.16 (.31) |
| Age (years) | 1.21 (.08) | 0.99 (.91) | 1.15 (.17) | 1.31 (.01) | 0.98 (.89) | 0.98 (.85) | 1.03 (.83) | 1.13 (.31) |
| Gender (female; reference) | 2.35 (<.009) | 2.00 (<.012) | 1.48 (<.018) | 2.15 (<.011) | 2.42 (<.007) | 2.43 (<.007) | 2.16 (<.011) | 2.11 (<.011) |

[a]Sell/distribute illegal drugs.

[b]Smoke cigarettes.

[c]Drink beer, wine, or hard liquor (eg, vodka, whiskey, or gin) when my parents do not know about it.

[d]Use marijuana.

[e]Use LSD, cocaine, amphetamines, or other illegal drugs.

[f]Use heroin.

[g]Use synthetic drugs (eg, K2, spice, bath salts).

[h]Use any prescription pills without a prescription.

[i]LTE: Living the Example.

# Discussion

## Principal Findings

LTE had a positive effect on reducing substance use intentions among high school students. Overall, results of this study were consistent with and indicate a stronger LTE effect in this study compared with the 2016-2017 previous pilot study [42]. LTE appears to offer a protective effect, with exposure to program messages leading to reduced drug use intentions. This approach deserves further development, refinement, and evaluation using rigorous randomized controlled methods that capture detailed dose-response data.

With regard to our hypotheses, first we saw some evidence that *ambassadors* were successful in reaching their peers. We were able to independently record social media posts for 24 *ambassadors* who produced LTE messages and shared them with nearly 13,000 followers over a period of 8 months. This demonstrates the feasibility of the LTE model reaching a large number of youths using a peer-to-peer approach at a low cost. Although the LTE program intended to leverage social media platforms that were most widely used by youth participants and their peer networks, some platforms, including Instagram, presented unique challenges for program implementers and evaluators. Instagram stories disappear from a user's profile after 24 hours, introducing barriers for documenting the content, features, and persuasive tactics used by *ambassadors*. Additionally, some ambassadors had private Instagram profiles, which prevented our documenting their posts. The abbreviated postduration also limits the extent to which the content can be viewed by followers and potentially shared on follower profiles. Regardless, LTE demonstrated significant potential to reach large numbers of peers via social media networks, with some months having over 10,000 followers of *ambassadors* on Instagram when there were only 14 *ambassadors* making at least one post during that period.

Second, we found that exposure to LTE is indeed associated with reduced 30-day drug use intentions, confirming our previous pilot study findings [42]. In our first set of regressions, we found that self-reported exposure to LTE was associated with lower use intentions across all 8 measured substances, including widely used drugs such as marijuana and prescription painkillers (opioids). This suggests that LTE is an effective approach that deserves large-scale implementation and evaluation.

Third, we found evidence of a dose-response relationship between LTE exposure and reduced 30-day drug use intentions. There was an effect of exposure to peer-to-peer messaging over time, with youths who self-reported exposure showing reduced use intentions at the wave 3 endline follow-up. Dose-response effects are important evidence for the effectiveness of media interventions, first as they demonstrate a direct relationship between the media stimulus and outcomes of interest and second as they provide evidence for the magnitude of those effects [62]. Optimizing dosage is an important objective for health campaigns [63] and for this research.

## Future Research and Limitations

Future research should use a randomized controlled trial design to compare implementing and nonimplementing schools and examine dose-response effects in detail. A group randomized trial involving sufficient numbers of schools for statistical power purposes assigned both to implement and not implement LTE across a school year would provide confirmatory evidence of the program's effectiveness. Together with detailed measurement of social media activity both through independent monitoring of posting activity at the individual *ambassador* level and through self-report, a future study such as this could specify the optimal dosage for LTE to produce meaningful prevention effects.

There are some limitations to this study. First, the number of survey respondents declined in wave 2 and 3 follow-ups, but we cannot confirm whether students were not in attendance during survey administration or they did not agree to complete the follow-up surveys. Second, we were only able to independently confirm LTE social media activity for 24 of the

46 *ambassadors*. We were unable to track potential posting for some *ambassadors* due to the omission of program hashtags in posts, due to use of the hashtags but with private Instagram account settings, or due to posting on the Snapchat platform, which does not permit hashtag tracking and is otherwise quite difficult to monitor. Additionally, we cannot quantify how many of the posts we documented, and engagement statistics we captured, represent unique individual users. Consequently, our independent social media data are incomplete. Finally, due to limitations with the social media monitoring, we were not able to statistically analyze the relationship between independent social media metrics and outcomes measured in the surveys.

## Conclusions

Adolescent substance use continues to be a pressing public health threat in the United States, in particular due to the growing availability of lethal drugs, teen social norms shifting to be more favorable of recreational use, and growing perceptions of the relative safety of some substances [2-4]. Digital social networks are major drivers of public perceptions of substance use [7,25,26], calling for novel digital strategies for prevention in this space. LTE offers a promising model for future programming seeking to widely disseminate peer-to-peer prevention messaging via social media to reduce adolescent drug use intentions. By training a relatively small number of youths, there is potential to reach a large audience with authentic, targeted prevention messages, given the widespread use of social networking platforms. Future research should build on the evidence of the effectiveness of LTE by implementing and evaluating the program at a larger scale using randomized controlled methods.

## Conflicts of Interest

None declared.

## References


1. Johnston LD, O'Malley P, Miech R, Bachman J, Schulenberg J. Deep Blue. Ann Arbor: Institute for Social Research, The University of Michigan; 2015. Monitoring the Future national survey results on drug use, 1975-2014: Overview, key findings on adolescent drug use URL: https://tinyurl.com/y76qlqqd [accessed 2020-03-09]
2. Keyes KM, Schulenberg J, O'Malley P, Johnston L, Bachman J, Li G, et al. The social norms of birth cohorts and adolescent marijuana use in the United States, 1976-2007. Addiction 2011 Oct;106(10):1790-1800 [FREE Full text] [doi: 10.1111/j.1360-0443.2011.03485.x] [Medline: 21545669]
3. Substance Abuse and Mental Health Services Administration. Washington, DC: Department of Health and Human Services; 2017. Key Substance Use and Mental Health Indicators in the United States: Results from the 2016 National Survey on Drug Use and Health URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-FFR1-2016/NSDUH-FFR1-2016.htm [accessed 2020-03-09]
4. Johnston LD, Miech R, O'Malley P, Bachman J, Schulenberg J, Patrick M. Deep Blue. Ann Arbor, MI: University of Michigan; 2019. Monitoring the Future national survey results on drug use, 1975-2018: Overview, key findings on adolescent drug use URL: https://tinyurl.com/ycqqoqet [accessed 2020-03-09]
5. Levy S. Youth and the Opioid Epidemic. Pediatrics 2019 Feb;143(2):pii: e20182752 [FREE Full text] [doi: 10.1542/peds.2018-2752] [Medline: 30602544]
6. Committee on Communications, American Academy of Pediatrics, Strasburger VC. Children, adolescents, and advertising. Pediatrics 2006 Dec;118(6):2563-2569. [doi: 10.1542/peds.2006-2698] [Medline: 17142547]
7. Miech RA, Johnston L, O'Malley PM, Bachman JG, Schulenberg J, Patrick ME. Trends in use of marijuana and attitudes toward marijuana among youth before and after decriminalization: the case of California 2007-2013. Int J Drug Policy 2015 Apr;26(4):336-344 [FREE Full text] [doi: 10.1016/j.drugpo.2015.01.009] [Medline: 25662893]
8. Schmidt LA, Jacobs LM, Spetz J. Young people's more permissive views about marijuana: local impact of state laws or national trend? Am J Public Health 2016;106(8):1498-1503. [doi: 10.2105/ajph.2016.303153]
9. Johnston LD, O'Malley P, Bachman J, Schulenberg J, Patrick M. Education Resources Information Center. Ann Arbor, MI: Institute for Social Research, The University of Michigan; 2017. Monitoring the Future National Survey Results on Drug Use, 1975-2016: Overview, Key Findings on Adolescent Drug Use URL: https://eric.ed.gov/?id=ED578534 [accessed 2020-03-09]
10. Curtin SC, Tejada-Vera B, Warmer M. Drug overdose deaths among adolescents aged 15-19 in the United States: 1999-2015. NCHS Data Brief 2017 Aug(282):1-8 [FREE Full text] [Medline: 29155681]
11. Miller T, Hendrie D. Substance Abuse and Mental Health Services Administration. 2008. Substance Abuse Prevention Dollars and Cents: A Cost-Benefit Analysis URL: https://www.samhsa.gov/sites/default/files/cost-benefits-prevention.pdf [accessed 2020-03-09]
12. Evans WD, Blitstein J, Hersey JC, Renaud J, Yaroch AL. Systematic review of public health branding. J Health Commun 2008 Dec;13(8):721-741. [doi: 10.1080/10810730802487364] [Medline: 19051110]
13. Patrick ME, Schulenberg JE, O'Malley PM, Johnston LD, Bachman JG. Adolescents' reported reasons for alcohol and marijuana use as predictors of substance use and problems in adulthood. J Stud Alcohol Drugs 2011 Jan;72(1):106-116 [FREE Full text] [doi: 10.15288/jsad.2011.72.106] [Medline: 21138717]

14. Lisdahl KM, Gilbart ER, Wright NE, Shollenbarger S. Dare to delay? The impacts of adolescent alcohol and marijuana use onset on cognition, brain structure, and function. Front Psychiatry 2013;4:53 [FREE Full text] [doi: 10.3389/fpsyt.2013.00053] [Medline: 23847550]
15. Centers for Disease Control and Prevention. Atlanta: US Department of Health and Human Services; 2014. Best Practices for Comprehensive Tobacco Control Programs URL: https://www.cdc.gov/tobacco/stateandcommunity/best_practices/pdfs/2014/comprehensive.pdf [accessed 2020-03-09]
16. Jager G, Block RI, Luijten M, Ramsey NF. Cannabis use and memory brain function in adolescent boys: a cross-sectional multicenter functional magnetic resonance imaging study. J Am Acad Child Adolesc Psychiatry 2010 Jun;49(6):561-72, 572.e1 [FREE Full text] [doi: 10.1016/j.jaac.2010.02.001] [Medline: 20494266]
17. Medina KL, Schweinsburg AD, Cohen-Zion M, Nagel BJ, Tapert SF. Effects of alcohol and combined marijuana and alcohol use during adolescence on hippocampal volume and asymmetry. Neurotoxicol Teratol 2007;29(1):141-152 [FREE Full text] [doi: 10.1016/j.ntt.2006.10.010] [Medline: 17169528]
18. McGuinness TM. Update on marijuana. J Psychosoc Nurs Ment Health Serv 2009 Oct;47(10):19-22. [doi: 10.3928/02793695-20090902-03] [Medline: 19835316]
19. Pedersen ER, Miles JN, Ewing BA, Shih RA, Tucker JS, D'Amico EJ. A longitudinal examination of alcohol, marijuana, and cigarette perceived norms among middle school adolescents. Drug Alcohol Depend 2013 Dec 1;133(2):647-653 [FREE Full text] [doi: 10.1016/j.drugalcdep.2013.08.008] [Medline: 24012070]
20. Walker DD, Neighbors C, Rodriguez LM, Stephens RS, Roffman RA. Social norms and self-efficacy among heavy using adolescent marijuana smokers. Psychol Addict Behav 2011 Dec;25(4):727-732 [FREE Full text] [doi: 10.1037/a0024958] [Medline: 21842969]
21. Friese B, Grube JW. Legalization of medical marijuana and marijuana use among youths. Drugs (Abingdon Engl) 2013 Feb 1;20(1):33-39 [FREE Full text] [doi: 10.3109/09687637.2012.713408] [Medline: 23641127]
22. Mason WA, Hanson K, Fleming CB, Ringle JL, Haggerty KP. Washington State recreational marijuana legalization: parent and adolescent perceptions, knowledge, and discussions in a sample of low-income families. Subst Use Misuse 2015 Apr;50(5):541-545 [FREE Full text] [doi: 10.3109/10826084.2014.952447] [Medline: 25671633]
23. NIDA for Teens. Bethesda, MD: NIDA Prescription Pain Medications (Opioids) URL: https://teens.drugabuse.gov/drug-facts/prescription-pain-medications-opioids [accessed 2020-01-31]
24. Jones CM, Einstein EB, Compton WM. Changes in synthetic opioid involvement in drug overdose deaths in the United States, 2010-2016. J Am Med Assoc 2018 May 1;319(17):1819-1821 [FREE Full text] [doi: 10.1001/jama.2018.2844] [Medline: 29715347]
25. Cerdá M, Wall M, Keyes KM, Galea S, Hasin D. Medical marijuana laws in 50 states: investigating the relationship between state legalization of medical marijuana and marijuana use, abuse and dependence. Drug Alcohol Depend 2012 Jan 1;120(1-3):22-27 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.06.011] [Medline: 22099393]
26. Schuermeyer J, Salomonsen-Sautel S, Price RK, Balan S, Thurstone C, Min S, et al. Temporal trends in marijuana attitudes, availability and use in Colorado compared to non-medical marijuana states: 2003-11. Drug Alcohol Depend 2014 Jul 1;140:145-155 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.04.016] [Medline: 24837585]
27. Greene K, Catona D, Elek E, Magsamen-Conrad K, Banerjee SC, Hecht ML. Improving prevention curricula: lessons learned through formative research on the youth message development curriculum. J Health Commun 2016 Oct;21(10):1071-1078 [FREE Full text] [doi: 10.1080/10810730.2016.1222029] [Medline: 27684111]
28. Slater MD, Kelly KJ, Lawrence FR, Stanley LR, Comello ML. Assessing media campaigns linking marijuana non-use with autonomy and aspirations: 'Be Under Your Own Influence' and ONDCP's 'Above the Influence'. Prev Sci 2011 Mar;12(1):12-22 [FREE Full text] [doi: 10.1007/s11121-010-0194-1] [Medline: 21271357]
29. Carpenter CS, Pechmann C. Exposure to the above the influence antidrug advertisements and adolescent marijuana use in the United States, 2006–2008. Am J Public Health 2011 May;101(5):948-954. [doi: 10.2105/ajph.2010.300040]
30. Evans WD, Holtz K, White T, Snider J. Effects of the above the influence brand on adolescent drug use prevention normative beliefs. J Health Commun 2014;19(6):721-737. [doi: 10.1080/10810730.2013.837559] [Medline: 24499473]
31. Hecht ML, Lee JK. Branding through cultural grounding: The keepin' it REAL curriculum. In: Evans DW, Hastings G, editors. Public Health Branding: Applying Marketing for Social Change. London: Oxford University Press; 2008.
32. Evans WD, Price S, Blahut S. Evaluating the truth brand. J Health Commun 2005 Mar;10(2):181-192. [doi: 10.1080/10810730590915137] [Medline: 15804907]
33. Evans WD, Taruberekera N, Longfield K, Snider J. Brand equity and willingness to pay for condoms in Zimbabwe. Reprod Health 2011 Oct 26;8:29 [FREE Full text] [doi: 10.1186/1742-4755-8-29] [Medline: 22029874]
34. Mentor Foundation USA. 2019 Apr 15. 'Shatter the Myths' Youth Rally at Glen Burnie High School URL: http://mentorinternational.org/usa%20on%20May%2029 [accessed 2020-01-30]
35. Substance Abuse and Mental Health Services Administration. Washington, DC: Department of Health and Human Services Results from the 2011 National Survey on Drug Use and Health: Summary of National Findings URL: https://tinyurl.com/ycbn2a9q [accessed 2020-03-09]
36. NIDA for Teens. 2014. Marijuana URL: http://teens.drugabuse.gov/drug-facts/marijuana [accessed 2019-06-02]

37. Jacobus J, Squeglia LM, Sorg SF, Nguyen-Louie TT, Tapert SF. Cortical thickness and neurocognition in adolescent marijuana and alcohol users following 28 days of monitored abstinence. J Stud Alcohol Drugs 2014 Sep;75(5):729-743 [FREE Full text] [doi: 10.15288/jsad.2014.75.729] [Medline: 25208190]
38. Evans WD, Rath JM, Hair EC, Snider JW, Pitzer L, Greenberg M, et al. Effects of the truth FinishIt brand on tobacco outcomes. Prev Med Rep 2018 Mar;9:6-11 [FREE Full text] [doi: 10.1016/j.pmedr.2017.11.008] [Medline: 29276667]
39. Calkins T. The challenge of branding. In: Calkins T, Tybout A, editors. Kellogg on Branding. New York: John Wiley and Sons; 2005.
40. Shek DT, Dou D, Zhu X, Chai W. Positive youth development: current perspectives. Adolesc Health Med Ther 2019;10:131-141 [FREE Full text] [doi: 10.2147/AHMT.S179946] [Medline: 31572041]
41. Evans WD, Andrade EL, Barrett ND, Cleary SD, Snider J, Edberg M. The mediating effect of Adelante brand equity on Latino immigrant positive youth development outcomes. J Health Commun 2018;23(7):606-613 [FREE Full text] [doi: 10.1080/10810730.2018.1496205] [Medline: 30138045]
42. Evans W, Andrade E, Goldmeer S, Smith M, Snider J, Girardo G. The living the example social media substance use prevention program: a pilot evaluation. JMIR Ment Health 2017 Jun 27;4(2):e24 [FREE Full text] [doi: 10.2196/mental.7839] [Medline: 28655704]
43. Pew Research Center. 2014 Apr 2. Section 2: Views of Marijuana – Legalization, Decriminalization, Concerns URL: http://www.people-press.org/2014/04/02/section-2-views-of-marijuana-legalization-decriminalization-concerns/ [accessed 2020-01-31]
44. Pew Research Center. 2014 Apr 2. Feds May Be Rethinking the Drug War, but States Have Been Leading the Way URL: http://www.pewresearch.org/fact-tank/2014/04/02/feds-may-be-rethinking-the-drug-war-but-states-have-been-leading-the-way/ [accessed 2019-05-26]
45. Neighbors C, Foster DW, Walker DD, Kilmer JR, Lee CM. Social identity as a moderator of the association between perceived norms and marijuana use. J Stud Alcohol Drugs 2013 May;74(3):479-483 [FREE Full text] [doi: 10.15288/jsad.2013.74.479] [Medline: 23490578]
46. Above the Influence. Bethesda, MD: NIDA Drug Facts URL: https://abovetheinfluence.com/drug-facts/ [accessed 2020-01-31]
47. Elliott S. The New York Times. 2013 Jun 15. Antidrug Campaign, Lacking Federal Funds, Turns to Social Media URL: http://www.nytimes.com/2013/07/16/business/media/antidrug-campaign-lacking-federal-funds-turns-to-social-media.html [accessed 2019-01-30]
48. Rothman AJ, Bartels RD, Wlaschin J, Salovey P. The strategic use of gain- and loss-framed messages to promote healthy behavior: how theory can inform practice. J Commun 2006;56(1):S202-S220. [doi: 10.1111/j.1460-2466.2006.00290.x]
49. Rothman AJ, Salovey P, Antone C, Keough K, Martin CD. The influence of message framing on intentions to perform health behaviors. J Exp Soc Psychol 1993;29(5):408-433. [doi: 10.1006/jesp.1993.1019]
50. O'Keefe DJ, Jensen JD. The relative persuasiveness of gain-framed and loss-framed messages for encouraging disease prevention behaviors: a meta-analytic review. J Health Commun 2007;12(7):623-644. [doi: 10.1080/10810730701615198] [Medline: 17934940]
51. Huhman ME, Potter LD, Nolin MJ, Piesse A, Judkins DR, Banspach SW, et al. The Influence of the VERB campaign on children's physical activity in 2002 to 2006. Am J Public Health 2010 Apr;100(4):638-645. [doi: 10.2105/AJPH.2008.142968] [Medline: 19608963]
52. Farrelly MC, Davis KC, Haviland ML, Messeri P, Healton CG. Evidence of a dose—response relationship between 'Truth' antismoking ads and youth smoking prevalence. Am J Pub Health 2005;95(3):425-431. [doi: 10.2105/ajph.2004.049692]
53. Evans WD, Price S, Blahut S, Hersey J, Niederdeppe J, Ray S. Social imagery, tobacco independence, and the TruthSM campaign. J Health Commun 2004 Sep;9(5):425-441. [doi: 10.1080/10810730490504413]
54. Mays D, Niaura RS, Evans WD, Hammond D, Luta G, Tercyak KP. Cigarette packaging and health warnings: the impact of plain packaging and message framing on young smokers. Tob Control 2015 Mar;24(e1):e87-e92 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051234] [Medline: 24420310]
55. Evans WD, Davis KC, Umanzor C, Patel K, Khan M. Evaluation of sexual communication message strategies. Reprod Health 2011 May 20;8:15 [FREE Full text] [doi: 10.1186/1742-4755-8-15] [Medline: 21599875]
56. McGuire W. Theoretical foundations of campaigns. In: Rice R, Atkin CK, editors. Public Communication Campaigns. Thousand Oaks, CA: Sage Publications; 1981.
57. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 Jan 30;211(4481):453-458. [doi: 10.1126/science.7455683] [Medline: 7455683]
58. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
59. Evans WD. Branding social and health behavior: An education and research agenda. In: Evans WD, editor. Psychology of Branding. Hauppauge, NY: Nova Science Publishers; 2013:133-148.
60. Evans W, Rath J, Pitzer L, Hair E, Snider J, Cantrell J, et al. Design and feasibility testing of the truth FinishIt tobacco countermarketing brand equity scale. J Health Commun 2016 Jul;21(7):800-808. [doi: 10.1080/10810730.2016.1157658] [Medline: 27315354]

61. National Institute on Drug Abuse. Monitoring the Future URL: https://www.drugabuse.gov/related-topics/trends-statistics/monitoring-future [accessed 2020-01-30]
62. Davis KC, Alexander RL, Shafer P, Mann N, Malarcher A, Zhang L. The dose-response relationship between tobacco education advertising and calls to Quitlines in the United States, March-June, 2012. Prev Chronic Dis 2015 Nov 5;12:E191 [FREE Full text] [doi: 10.5888/pcd12.150157] [Medline: 26542143]
63. Abroms LC, Ahuja M, Kodl Y, Thaweethai L, Sims J, Winickoff JP, et al. Text2Quit: results from a pilot test of a personalized, interactive mobile health smoking cessation program. J Health Commun 2012;17(Suppl 1):44-53 [FREE Full text] [doi: 10.1080/10810730.2011.649159] [Medline: 22548598]

## Abbreviations

**GW:** George Washington
**LSD:** lysergic acid diethylamide
**LTE:** Living the Example
**NIDA:** National Institute on Drug Abuse
**PYD:** Positive Youth Development
**QED:** quasi-experimental design
**SAMHSA:** Substance Abuse and Mental Health Services Administration
**STM:** Shatter the Myths

*Edited by G Eysenbach; submitted 10.09.19; peer-reviewed by J Toelen, HJ Choi, E Hutson; comments to author 21.01.20; revised version received 02.02.20; accepted 22.02.20; published 06.05.20.*

*Please cite as:*
*Evans W, Andrade E, Pratt M, Mottern A, Chavez S, Calzetta-Raymond A, Gu J*
*Peer-to-Peer Social Media as an Effective Prevention Strategy: Quasi-Experimental Evaluation*
*JMIR Mhealth Uhealth 2020;8(5):e16207*
*URL: https://mhealth.jmir.org/2020/5/e16207*
*doi:10.2196/16207*
*PMID:32374270*

Original Paper

# A Sexual Health Promotion App for Transgender Women (Trans Women Connected): Development and Usability Study

Christina J Sun[1], PhD, MS; Kirsten M Anderson[1], MSW, MPH; Tamara Kuhn[2], MA; Liat Mayer[3], MA; Charles H Klein[3], JD, PhD

[1]School of Public Health, Oregon Health & Science University-Portland State University, Portland, OR, United States

[2]dfusion, Scotts Valley, CA, United States

[3]Department of Anthropology, Portland State University, Portland, OR, United States

**Corresponding Author:**
Christina J Sun, PhD, MS
School of Public Health
Oregon Health & Science University-Portland State University
506 SW Mill St., Suite 450
SCH
Portland, OR, 97201
United States
Phone: 1 503 725 8858
Email: chrsun@pdx.edu

## Abstract

**Background:** HIV severely impacts the transgender communities in the United States, and transgender women have the highest HIV incidence rates among any identified risk group. Guided by formative research with transgender women and by an expert advisory panel of transgender women, we designed a prototype mobile app to promote HIV prevention among transgender women.

**Objective:** This study aimed to develop and test the usability and acceptability of the prototype Trans Women Connected mobile app.

**Methods:** We engaged in a 3-phase prototype development process. After conducting formative research about the health needs of this population, we outlined a theory-based app framework and developed three prototype activities (ie, a vision board, a pre-exposure prophylaxis [PrEP] education activity, and an interactive map). We then tested the usability and acceptability of the mobile app and activities with 16 transgender women using pre- and posttests, think-aloud protocols, and open-ended questions.

**Results:** Participants reported high acceptability for the mobile app; the mean rating across all usability and likability questions was 5.9 out of 7. Service utilization intention, goal setting, and social support increased at posttest compared with pretest. Increases in self-efficacy in finding lesbian, gay, bisexual, transgender, and queer–friendly services; intention to seek online social support; and PrEP knowledge were statistically significant. Participants described the app as attractive and useful and perceived all three activities positively.

**Conclusions:** This study describes the development and usability and acceptability evaluation of a prototype mobile app designed for and with transgender women for HIV prevention. The usability testing findings provided important insights toward refining and the further development of the Trans Women Connected mobile app. The results suggest that a mobile health intervention can support positive changes. The remaining development and efficacy randomized trial of the Trans Women Connected mobile app is currently underway.

***(JMIR Mhealth Uhealth 2020;8(5):e15888)*** doi:10.2196/15888

**KEYWORDS**

transgender persons; HIV; sexual health; minority health; health care disparities; health status disparities

## Introduction

According to a 2016 meta-analysis, nearly 1 million adults in the United States are estimated to be transgender [1]. Transgender (or trans, in short) is a term for individuals whose gender expression and/or gender identity do not align with cultural expectations and gender norms associated with their sex assignment at birth [2]. Transgender women in the United States have been severely impacted by HIV [3-8], and more than 25% of them are living with HIV [5]. Transgender women of color are particularly affected; 56% of black/African American transgender women are living with HIV [5]. HIV incidence rates among transgender people are much higher than the national average [9] and highest among any group specifically tracked by the Centers for Disease Control and Prevention (CDC) [10]. Currently, the only transgender-specific program in the CDC's Effective Interventions and Compendium of Evidence-Based Interventions and Best Practices for HIV [11,12] is an in-person, couples-based intervention designed for transgender women coupled with a cisgender male partner [13].

Internet/electronic health (eHealth) and mobile health (mHealth) modalities may be particularly appropriate for transgender women, many of whom are socially marginalized, who live in areas that do not offer transgender-specific programming and/or may be uncomfortable participating in face-to-face interventions because of confidentiality concerns and fear of stigmatization [5,14-16]. Although empirical data on internet and mobile app use among transgender persons is limited, existing studies show that transgender women consistently use their phones for information gathering, socializing, and making sexual connections [17-19]. These communication practices provide opportunities for engaging transgender women in HIV prevention and sexual health promotion issues. Mobile phone apps also have the capacity to deliver individually tailored content based on motivations for use and app use patterns [20], making them particularly appropriate for addressing the heterogeneous experiences, identities, and needs of transgender women across the life course. Furthermore, although there are very few studies evaluating transgender-specific eHealth or mHealth programs [21], a growing body of research about diverse communities shows that new media intervention programs are acceptable and can be effective in reducing HIV/sexually transmitted infection–related risk behaviors and linking individuals to prevention and care services [22-34]. Given these realities, program developers and practitioners are increasingly calling for transgender-specific internet-, social media-, and mobile-based programs to expand the reach of HIV prevention and health promotion activities [3,35-38].

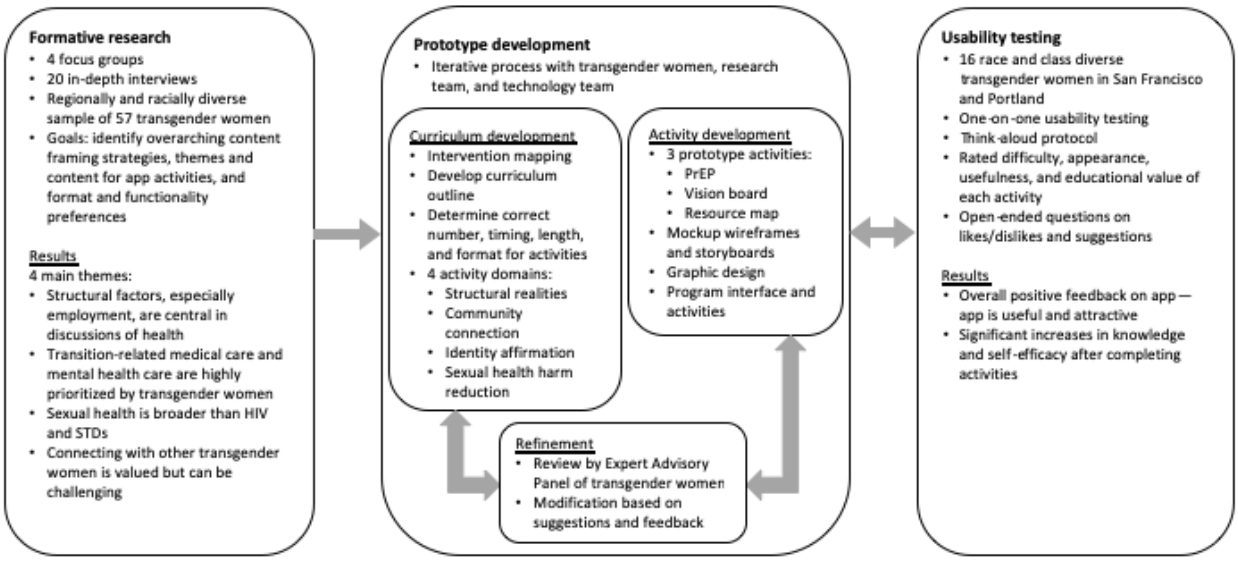
The purpose of this study was to develop and evaluate the usability and acceptability of a prototype of the Trans Women Connected mobile app that addresses the unique HIV prevention and sexual health needs of transgender women. The goal was to engage transgender women through a strengths-based approach to HIV prevention and sexual health promotion, leveraging the power of social networks to identify and encourage protective factors.

## Methods

### Overview

We developed the prototype in three phases (see Figure 1). This process occurred in collaboration with a research team, an expert advisory panel (EAP), and a technology team. The EAP consisted of 3 transgender women of color, a Latinx woman, a Filipina woman, and an African American woman, each of whom holds a high-level position in an organization providing services to transgender women. They provided consultation and feedback throughout the prototype design and development process and were compensated as project consultants. The technology team included a technology director, an app developer, a database engineer, and a user experience and graphic designer and was responsible for the design and programming of the prototype. The ETR Institutional Review Board approved the study protocols and provided oversight. In phase one, the research team conducted formative research with potential users of the mobile app to understand the health needs of this population and how an app might support their health and to gather specific suggestions for the development and content of the app. Phase two included prototype design and development of the overall app structure and three activities (ie, a vision board, a pre-exposure prophylaxis [PrEP] education activity, and an interactive map). In phase three, a group of potential users used, rated the usability and acceptability of, and described their experiences with the prototype.

**Figure 1.** Prototype development process. PrEP: pre-exposure prophylaxis; STD: sexually transmitted disease.




## Phase One: Formative Research

The research team conducted four focus groups and 20 in-depth qualitative interviews with transgender women between the ages of 18 and 59 years living in urban and rural areas in every region of the United States [39]. Briefly, women were recruited through partnerships with community-based organizations, fliers in community spaces and on social media, and word of mouth. Focus group participants received US $100 in compensation; interview participants received US $50. Topics included overall well-being and connectedness, transgender health, sexual health, use of internet and social media, and recommendations for a trans-specific app. Focus groups and interviews were audio-recorded and transcribed. We used the grounded theory open coding methodology and ethnographic methodologies to classify key themes [40,41].

## Phase Two: Prototype Design and Development

On the basis of the findings of the formative research, the research team developed a preliminary curriculum outline, which was discussed with and refined by the EAP and, at times, additional transgender women from their networks. The technology team then determined the optimal formats for conveying the curriculum in an engaging manner, in small self-paced segments, with sufficient levels of interactivity, and customization. The technology team next created wireframes and storyboards to mock-up the layout, workflow, and interaction of all app elements. Through an iterative process of review and feedback, the EAP guided the designer to create a final set of imagery and style guide (see Figure 2 for the overall look). Following final approval of the storyboards, the app developer created the prototype framework using Adobe Animate. Using this cross-platform development method, the app could be built for several mobile platforms, including Android and iOS, and would allow the utilization of integrated mobile device (eg, mobile phone and tablet) features such as GPS and camera.

**Figure 2.** Welcome screen for the Trans Women Connected mobile app.


### Phase Three: Usability Testing

#### *Participants*

Two community-based organizations, one in San Francisco, California and one in Portland, Oregon, recruited participants. Inclusion criteria were ages 18 to 55 years, self-identify as a transgender woman, and English speaking. All participants provided written informed consent. Participants received US $100 in compensation.

#### *Procedures and Measures*

Using a standard *think-aloud* protocol, the usability test administrator (TK) asked each participant to complete a series of tasks associated with typical use case scenarios, such as using the menu to locate activities and completing activities. We also objectively measured participants' task completion success and time to completion and asked them to articulate their questions and thinking process [42,43].

Participants rated the difficulty, appearance, perceived usefulness, and educational value following each key task and after completion of all activities and completed a pre- and posttest that assessed service utilization, goal setting, social support, and HIV PrEP knowledge. We also solicited open-ended comments regarding their perceptions of the app and suggestions to improve it.

#### *Analysis*

We conducted descriptive analyses of the ratings and assessed changes from pre- to posttest using nonparametric tests in IBM SPSS version 24 (IBM Corp, Armonk, NY). The research team reviewed open-ended comments and identified key suggestions.

## *Results*

### Phase One: Formative Research

#### *Themes About Transgender Health*

We identified four main themes from the focus groups and interviews with 57 transgender women (45/57, 79% who were transgender women of color) related to the health of transgender women:

1. Employment and structural factors play a key role in the health of transgender women.
2. Transition-related care and mental health care are highly prioritized by transgender women. However, knowledgeable and affirming providers are difficult to find.
3. The sexual health needs of transgender women are much broader than just HIV and sexually transmitted infections.
4. Finding community and social support is challenging, but an important part of health for transgender women.

#### *Conceptual Feedback About a Mobile App for Transgender Health*

Our formative research found that most participants regularly used social media channels such as Facebook, Instagram, Snapchat, and various social and sexual networking apps to connect with friends, family, romantic/sexual partners, and work opportunities. However, the women also believed that the Trans Women Connected mobile app could fill a gap by providing resources specific to transgender women and an opportunity to connect to other transgender women (see Textbox 1). One Atlanta focus group participant described how this app could be a tool for "lifting the curtain and letting go of baggage." In particular, participants valued possible app features, such as forums and direct messaging, to enable users to engage with each other and for the possibility of formal mentoring.

Participants had questions and concerns about how the app would be structured, including issues of privacy and safety, and the possibility of forums and messaging being used for dating or hooking up. In general, however, participants were positive about the mobile app, and many expressed that it would fulfill an important need in the community.

**Textbox 1.** Suggestions for Trans Women Connected mobile app features and activities.

- Give transgender women a voice in the process at every step of development
- Images and videos used in the app depict actual transgender women
- Space for trans women to connect and share information
- Optional and customizable notifications
- Customizable privacy settings to allow users to control how their information is displayed
- Information about sexual health, including sexual pleasure
- Support for healthy relationships and relationship skills
- Information to help women locate both trans-specific resources and trans friendly businesses and organizations
- Ways to block and report other users

## Phase Two: Prototype Development

Guided by the formative research and EAP, the technology team created three prototype interactive activities. The first activity, entitled Create Your Own Vision (Board), provides participants with the opportunity to imagine how they would like their lives to be in eight domains: health, living situation, school/work, sex/love, making a difference, people, free time, and spirituality (see Figure 3). This activity seeks to help transgender women situate their well-being and sexual health within the social determinants of health and structural factors that emerged as the dominant theme of the focus groups and interviews.

This interactive vision board allows participants to use their existing pictures, take pictures, create doodles/artwork, select quotes, and add their own text or other images to create a vision for each of the eight life domains. These eight collages are merged into a final scrollable collage that represents the user's vision for their future. The activity also has the user identify a primary goal for their life today, lay out the steps needed to reach that goal, and identify who can help them execute each step. This goal setting activity provides a roadmap to reaching their goal that users can return to anytime. In addition, the vision board can be printed or shared with other app users—an option that all usability testers said they would use.

The second activity—Is PrEP Right for Me or My Partners?—presents a basic overview of PrEP and guides participants through a benefits and risks assessment to help them decide if PrEP might be appropriate for them or their partners (see Figure 4). The activity primarily presents information through a series of videos featuring a charismatic African American transgender woman who is getting ready for and enjoying a night out with friends, including 2 transgender women who tell personal stories about using PrEP (see Figure 5). Interactive elements were included through questions and a poll that allows the user to see what other users are thinking about PrEP and a map that identifies PrEP providers in the users' area.

**Figure 3.** A screenshot of the Create Your Own Vision (Board) prototype activity.

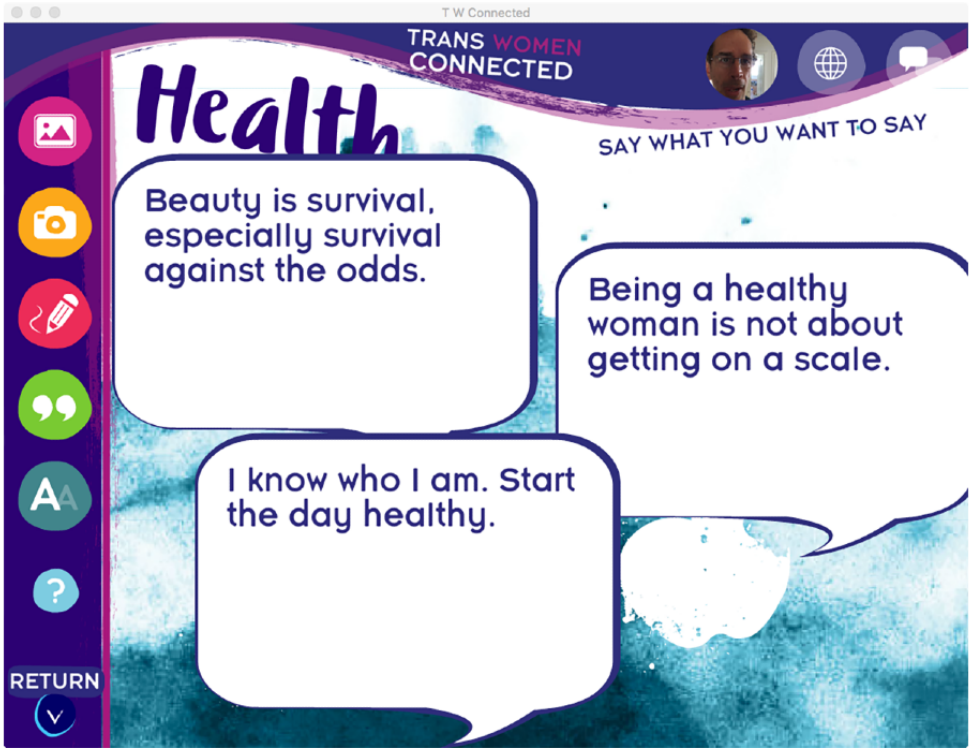



**Figure 4.** A screenshot of the Is pre-exposure prophylaxis Right for Me or My Partners? prototype activity.


Responding to the need for helping transgender women find culturally competent and welcoming services, the staff developed the third activity: Interactive Resource Map. The map displays providers/organizations in five service categories: medical, support, educational, employment, and housing/living. On the basis of GPS location, users see icons for each entry and can customize the map to view all or certain categories. When users choose an icon, they see details about the resource including name, contact information, and user ratings and reviews (see Figure 6). In addition, the map allows users to add locations to the map as they identify transgender-friendly providers in their area.

**Figure 5.** A screenshot of a video within the Is pre-exposure prophylaxis Right for Me or My Partners? prototype activity.

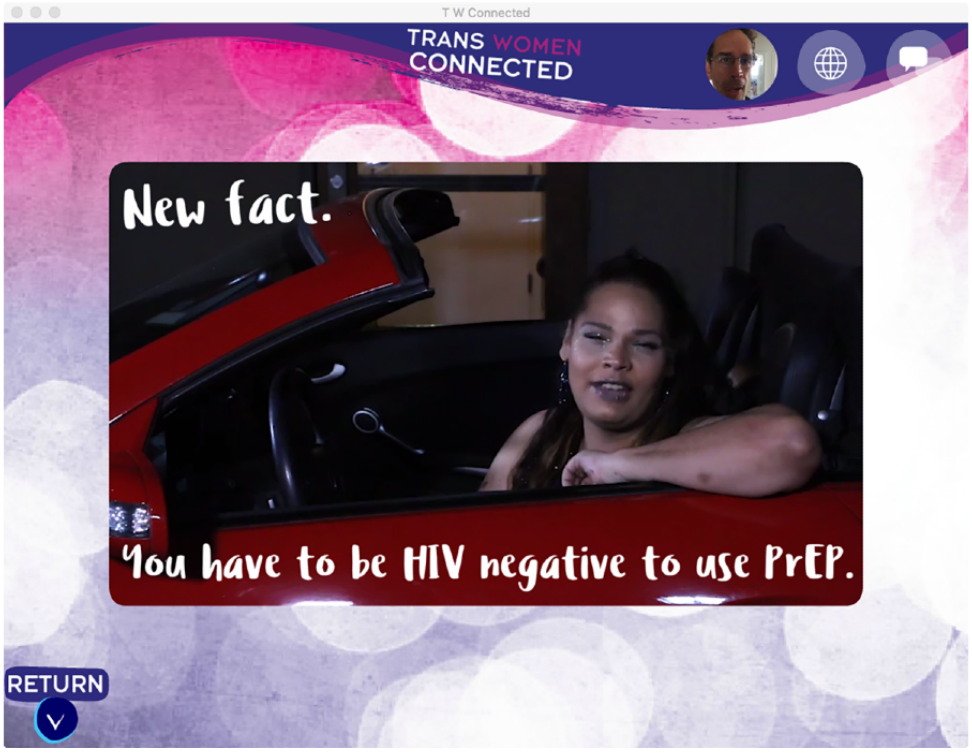
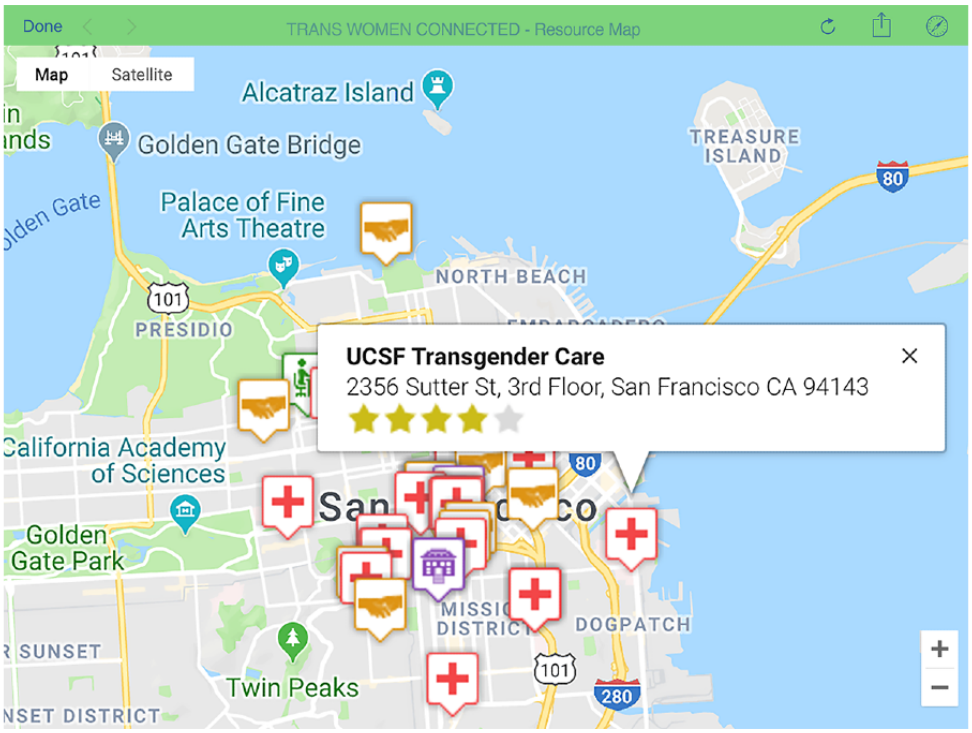



**Figure 6.** A screenshot of the Interactive Resource Map prototype activity.




## Phase Three: Usability Testing

### *Participant Characteristics*

A total of 16 racially diverse transgender women participated in the usability testing (see Table 1). Participants were aged 19 to 52 years, with an average of 34.5 years, and all were experienced with using mobile devices.

**Table 1.** Participant characteristics.

| Characteristics | Values |
|---|---|
| **Location, n (%)** | |
| San Francisco | 11 (69) |
| Portland | 5 (31) |
| **Gender identity, n (%)** | |
| Transgender woman | 14 (88) |
| Nonbinary | 2 (13) |
| **Race[a], n (%)** | |
| American Indian or Alaska Native | 3 (19) |
| Asian | 5 (31) |
| Black or African American | 3 (19) |
| Native Hawaiian or other Pacific Islander | 1 (6) |
| White/Caucasian | 6 (38) |
| Other | 3 (19) |
| **Ethnicity, n (%)** | |
| Hispanic or Latinx | 2 (13) |
| Non-Hispanic or Latinx | 14 (88) |
| Age (years), mean (SD) | 34.5 (9.28) |
| **Experience with mobile devices, n (%)** | |
| Very experienced | 14 (88) |
| A lot of experience | 1 (6) |
| Some experience | 1 (6) |
| Not much experience | 0 (0) |
| No experience | 0 (0) |
| **Primary mode to connect with transgender women, n (%)** | |
| Online | 0 (0) |
| In person | 7 (44) |
| Both online and in person | 8 (50) |
| Other | 1 (6) |
| **Primary mode to meet transgender women, n (%)** | |
| Online | 3 (19) |
| In person | 9 (56) |
| Both online and in person | 4 (25) |
| Other | 0 (0) |

[a]Participants could select more than one option.

### *App Usability*

The usability test results demonstrated that participants liked the mobile app and found it highly usable (see Table 2). The mean rating across all usability/likability questions was 5.9 out of 7. The study benchmark was a score of 4 or better for each individual rating component.

**Table 2.** Usability and likability ratings of the Trans Women Connected mobile app.

| Variable | Values, mean (SD)[a] |
|---|---|
| **Main menu** | |
| Likeability of opening screen | 6.25 (1.24) |
| Likeability of menu system | 5.66 (1.81) |
| **Create Your Own Vision (Board)** | |
| Likeability of the appearance | 6.28 (1.49) |
| Likeability of the interface | 6.44 (0.73) |
| Likeability of the narrator | 5.80 (1.14) |
| **Is PrEP[b] Right for Me or My Partners?** | |
| Likeability overall | 5.57 (1.60) |
| Likeability of women in the videos | 6.14 (1.03) |
| **Interactive Resource Map** | |
| Usefulness | 6.73 (0.59) |
| **Overall ratings of Trans Women Connected** | |
| Appearance | 6.38 (1.15) |
| Contents | 6.39 (0.65) |
| Ease of use | 6.25 (1.29) |
| Interest | 6.08 (1.16) |
| Professionally designed | 5.92 (1.38) |
| Enjoyability | 6.25 (1.48) |
| Usefulness | 6.50 (1.24) |
| Educational value | 6.25 (0.87) |
| Likelihood of using | 6.58 (0.68) |
| Likelihood of recommending | 6.62 (1.39) |

[a]7 represents the most positive or highest rating; the study benchmark was 4.

[b]PrEP: pre-exposure prophylaxis.

### *Changes in Service Utilization, Goal Setting, Social Support, and Pre-Exposure Prophylaxis Knowledge*

There was a positive trend and increase across multiple measures (eg, perceptions of available social support, service utilization intention, intention to mentor other transgender women on the app, and self-efficacy in discussing PrEP; see Table 3) after using the app. Increases in self-efficacy in finding lesbian, gay, bisexual, transgender, and queer (LGBTQ)–friendly services; intention to seek online social support; and PrEP knowledge were statistically significant.

**Table 3.** Self-efficacy, intention, knowledge, goal setting, and social support: pre– and post–usability testing scores.

| Variable | Usability testing scores | | *P* value |
|---|---|---|---|
| | Pre | Post | |
| Self-efficacy in finding LGBTQ[a]-friendly services, mean (SD) | 3.19 (0.75) | 3.69 (0.48) | .01 |
| Likelihood of seeking LGBTQ-friendly services, mean (SD) | 3.69 (0.48) | 3.56 (0.81) | .48 |
| Intention to make medical appointment[b], mean (SD) | 3.88 (0.34) | 3.88 (0.50) | >.99 |
| Created health goals, n (%) | 13 (81) | 13 (81) | >.99 |
| Considered steps to reach goal[c], n (%) | 9 (69) | 9 (69) | >.99 |
| Have social support to help reach goal, n (%) | 12 (75) | 14 (88) | .16 |
| Availability of social support, mean (SD) | 2.63 (1.02) | 3.00 (0.97) | .71 |
| Intention to seek online support[b], mean (SD) | 2.56 (1.09) | 3.06 (1.00) | .03 |
| Intention to seek local support services[b], mean (SD) | 3.38 (0.81) | 3.50 (0.63) | .48 |
| Helpfulness of mentoring from another trans woman, mean (SD) | 4.06 (1.24) | 4.19 (1.28) | .91 |
| Intention to mentor other transgender women on app, mean (SD) | 3.38 (0.89) | 3.56 (0.63) | .48 |
| Can identify friends/family for support, n (%) | 15 (94) | 16 (100) | >.99 |
| PrEP[d] knowledge, mean (SD) | 2.50 (1.37) | 3.56 (0.63) | .008 |
| Intention to seek more info about PrEP[b], mean (SD) | 2.80 (1.21) | 2.81 (1.22) | .58 |
| Intention to discuss PrEP with provider[b], mean (SD) | 2.07 (0.80) | 2.00 (0.97) | >.99 |
| Intention to discuss PrEP with partner[b,e], n (%) | 6 (40) | 6 (40) | >.99 |
| Self-efficacy in discussing PrEP, mean (SD) | 3.00 (1.13) | 3.25 (0.86) | .16 |
| Connection to other transgender women, mean (SD) | 6.44 (2.60) | 6.63 (2.33) | .78 |
| Satisfaction with connection to other transgender women, mean (SD) | 5.00 (1.67) | 5.19 (1.83) | .26 |
| Level of social support from other transgender women, mean (SD) | 5.19 (1.87) | 5.21 (1.81) | .72 |

[a]LGBTQ: lesbian, gay, bisexual, transgender, and queer.

[b]In the next 30 days.

[c]Out of 13 participants.

[d]PrEP: pre-exposure prophylaxis.

[e]Out of 15 participants.

### *Qualitative Findings From Usability Testing*

Usability testing participants also provided a number of qualitative comments about their experience using the prototype and suggestions for how it might be improved (see Textbox 2). Overall, this qualitative feedback was positive. As one participant told us, “Helping trans people navigate their health is a win.” As in the quantitative ratings, women expressed that they felt the app was useful and attractive. The welcome screen was described as being friendly and approachable, and many participants listed the esthetics of the app as one of their favorite parts. The women liked all three activities, with the map being a particular favorite. The potential for interactivity and connection with other transgender women was described as another attractive feature. The suggestions provided ranged from broad to very specific; a list of some of the most common or useful suggestions can be seen in Textbox 2.

**Textbox 2.** Suggestions to improve the Trans Women Connected mobile app prototype.

General

- More information or instructions on welcome screen and main menu
- Tidier appearance/colors
- More photos with a variety of transgender women (more representative of the community)
- Options for text instead of audio; subtitles for all audio
- Change or remove audio when returning to a screen
- More black trans women developers
- Provide suggestions for donating or volunteering

Create Your Own Vision (Board)

- Improve instructions and make navigation clearer
- Ability to draw directly on vision board
- Ability to add songs, gifs, or other "moving" content
- Option to turn off narration
- Option to share on social media

Is PrEP Right for Me or My Partners?

- Option to pause videos
- Shorter videos
- Option to view more scientific information
- Clarify some information

Interactive Resource Map

- Add a search bar and list of locations
- Ability to filter user-added locations
- Filter or information on Americans with Disabilities Act-accessible locations
- Add other types of resources, such as gender-neutral restrooms
- Add a filter or new category for free activities

# Discussion

## Principal Findings

We conducted a 3-phase process to develop a prototype HIV prevention mobile app for and with transgender women. In the first phase, we completed formative research, which indicated that HIV prevention for transgender women must address the larger context of women's lives. In the second phase, we developed and refined the overall app framework and three prototype activities based on feedback from the EAP. In the third phase, we conducted usability testing. Participants rated the app highly on usability and likeability. There was a significant increase in self-efficacy in finding LGBTQ-friendly services, intention to seek online social support, and PrEP knowledge. Overall, the participants were supportive of the app, with many saying they would use it regularly with no additional incentive beyond the information and opportunity to connect with other transgender women.

## Strengths of the Approach

There are a number of strengths of the Trans Women Connected prototype mobile app and research process. The 3-phase development process allowed us to collect formative data about the health priorities of transgender women and incorporate this information into the app prototype. We developed these activities because they addressed issues or requests observed in the formative research. This approach, which is considerably more extensive than the typical app development approach, helps to ensure that the selected prototype activities were addressing real needs of the community. The extensive formative research and the rigor of the analysis in this project, when combined with an iterative usability testing approach, are unique among mobile app developers.

Although researchers have called for digital health interventions to (1) be grounded in behavioral theory [44,45], (2) utilize an in-depth qualitative understanding of the intended population [46], and (3) be iteratively developed with multiple stages of user feedback [47,48], the process has been rarely implemented consistently, with variations in the development process between

academic researchers and technology industry developers. Most mHealth interventions designed by those within the industry do not yet incorporate theory-based strategies known to drive changes in health behaviors or undergo systematic testing to demonstrate their effectiveness [45,49,50]; however, they typically employ a rapid iterative development process of the technology. In contrast, interventions developed within an academic environment are more often grounded in behavioral theory [51], but researcher-driven mobile interventions often do not benefit from rapid iterative prototyping and standardized usability testing with iterative feedback [52].

The development of this app integrated the strategies used in both academia and the industry and, in some instances, used methodologies that were more rigorous. For example, rather than use a local convenience sample for focus groups, as is typical of most software and app developers, the project conducted focus groups with transgender women in four regions and interviews with transgender women throughout the United States, with key transgender community stakeholders actively assisting in recruitment activities. By using a community-driven, grounded theory approach, the research team was able to identify and delve deeply into themes that would not have been detected through simply examining focus group transcripts from convenience samples.

Furthermore, input from transgender women was integrated throughout the prototype development in coordination with the EAP and women from their networks. These iterative feedback loops began earlier than typically in app development and included gathering feedback on colors and image representations of transgender women and several other functional and aesthetic elements of the app before moving on to feedback on the app interface and activity content, function, and appearance. The inclusion of members of the LGBTQ community as part of the development process extended beyond formative research into voice talent/narration, video development, and design, with the designer being an LGBTQ-identified woman of color, the narrators being transgender women of color, and all video participants being transgender women. These multiple forms of community engagement not only resulted in a highly tailored app but also further provide a foundation for dissemination and active utilization once the Trans Women Connected mobile app is finalized. As suggested by other researchers, community engagement increases the likelihood the knowledge gained and the intervention, in this case an app, will benefit the community [53-55]. We recognize and appreciate the importance of the community's priorities and solutions, which enhances the relevance and use of data and quality and validity of research [56-58].

Additional strengths of the usability testing were the multiple types of data and an evaluation of impact. We gathered both quantitative and qualitative data about the experience of using the prototype mobile app; the quantitative data allowed us to verify that we were meeting numeric targets for usability and acceptability (as is typical for usability testing), whereas the qualitative data gave participants the opportunity to provide more in-depth feedback and suggestions. However, measuring the impact of the prototype on knowledge, attitudes, and intentions during the development process rather than when the app is completed goes considerably beyond standard usability testing, which is typically limited only to identifying problems within the user experience [59]. This approach helps determine whether the learning and engagement activities are performing as intended before building out the entire app.

Finally, a central strength of this mobile app is the emphasis on addressing HIV holistically using a strengths-based perspective. This approach recognizes the resilience of the transgender community and the social and structural barriers that transgender women face [60]. Developing an effective intervention requires a comprehensive approach that intentionally targets community strengths and challenges [60-63]. Our formative research confirmed that attention to the larger context of trans women's lives was key to HIV prevention [14], [61,64,65], and this understanding guided our development process.

## Limitations

A few limitations of our study need to be considered. The usability testing sample was limited to 16 participants in two different cities in the West Coast. Research has determined that 7 to 10 usability participants will uncover 80% to 90% of usability problems [66,67], and we were able to test the mobile app with a diverse group of 16 participants. This sample allowed the examination of some of the different experiences and identities of transgender women. We likely identified functional problems with the app and the app was well received as usable and acceptable; however, it is possible that more geographically diverse participants or participants from rural areas would have qualitatively experienced the app differently.

The inclusion of pre- and posttest knowledge and intention measures during a usability test is unusual and presents some limitations. Although significant increases in knowledge and intent were found on some variables, the sample size, although adequate for usability testing, is small for such testing and again may not represent the effect that would have been found with a more diverse or larger sample. However, the results are promising given that there were positive changes in key measures after viewing only three prototype activities, and this would seem to bode well for the efficacy of the complete mobile program. For several measures, there was no significant increase between pre- and posttest scores; we may have observed a ceiling effect. Given the short duration of app use, we were unable to measure behavior change. However, we measured antecedents of behavior change according to behavior change theories, and in the full trial when participants have a longer period to use the mobile app, we expect to observe increases in these antecedents and more protective behaviors.

A final limitation of the project was that all members of the research team identify as cisgender. Although the input of transgender women was included throughout the development process, no transgender individuals participated directly in the interpretation and analysis of data. As a result, there may be aspects of the data that were missed or misinterpreted because of the absence of this perspective.

## Next Steps

Given the positive results from the prototype development and usability testing, our next steps are to complete the development

of the mobile app, including the storyboards and curriculum/content for the remaining activities, and conduct a cluster randomized controlled trial to assess the effectiveness of the Trans Women Connected mobile app. The expansion of the app to include additional activities will be based on a theoretical framework of gender affirmation, resilience, and cognitive behavioral theory and will be implemented using an agile development process with iterative testing. Through the entire process, we will continue to work with an expanded expert and community advisory group of transgender women who will provide feedback on the app; they will review materials to ensure the content reflects their experiences and those of their community. The project team, in collaboration with the expert advisors, will identify the most essential topics, which will then be reviewed and revised through an iterative process of refinement. A newly formed community advisory board, consisting of 10 geographically and racially diverse transgender women, will provide feedback on the activity plans and storyboards. The project team will then revise the documents, with the process continuing until the content is approved by the community advisors. In addition, the community advisory board will test and review completed interactive activities throughout the project.

## Conclusions

Strong formative research, followed by an iterative review from members of the intended audience, resulted in a usable and acceptable mobile app that has been well received by transgender women. Usability testing findings provided important insights toward refining and the further development of the Trans Women Connected mobile app. The results of this study suggest that an mHealth intervention can address critical structural factors that shape transgender women's lives and support positive changes in the knowledge and attitudes about social connection and health access that are antecedents to increasing protective and health promotion behaviors.

## Acknowledgments

The authors would like to extend their gratitude for the time and effort of the transgender women who provided input and feedback on their app in the formative research interviews and focus groups, as members of the EAP, and in usability testing. They would also like to thank the community-based organizations that supported this project. This project was supported by grant R43MD012279 from the National Institute on Minority Health and Health Disparities and K12HS022981 from the Agency for Healthcare Research and Quality. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Minority Health and Health Disparities and the Agency for Healthcare Research and Quality.

## Conflicts of Interest

None declared.

## References

1. Institute of Medicine, Board on the Health of Select Populations, Transgender Health Issues and Research Gaps and Opportunities Committee on Lesbian, Gay, Bisexual. The Health Of Lesbian, Gay, Bisexual, And Transgender People: Building A Foundation For Better Understanding. Washington DC: The National Academies Press; 2011.
2. Meerwijk EL, Sevelius JM. Transgender population size in the United States: a meta-regression of population-based probability samples. Am J Public Health 2017 Feb;107(2):e1-e8. [doi: 10.2105/AJPH.2016.303578] [Medline: 28075632]
3. Centers for Disease Control and Prevention. 2019. HIV And Transgender People URL: http://www.cdc.gov/hiv/group/gender/transgender/ [accessed 2019-08-11]
4. Clements-Nolle K, Marx R, Guzman R, Katz M. HIV prevalence, risk behaviors, health care use, and mental health status of transgender persons: implications for public health intervention. Am J Public Health 2001 Jun;91(6):915-921. [doi: 10.2105/ajph.91.6.915] [Medline: 11392934]
5. Herbst JH, Jacobs ED, Finlayson TJ, McKleroy VS, Neumann MS, Crepaz N, HIV/AIDS Prevention Research Synthesis Team. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav 2008 Jan;12(1):1-17. [doi: 10.1007/s10461-007-9299-3] [Medline: 17694429]
6. Operario D, Soma T, Underhill K. Sex work and HIV status among transgender women: systematic review and meta-analysis. J Acquir Immune Defic Syndr 2008 May 1;48(1):97-103. [doi: 10.1097/QAI.0b013e31816e3971] [Medline: 18344875]
7. Poteat T, Scheim A, Xavier J, Reisner S, Baral S. Global epidemiology of HIV infection and related syndemics affecting transgender people. J Acquir Immune Defic Syndr 2016 Aug 15;72(Suppl 3):S210-S219 [FREE Full text] [doi: 10.1097/QAI.0000000000001087] [Medline: 27429185]
8. Reback CJ, Fletcher JB. HIV prevalence, substance use, and sexual risk behaviors among transgender women recruited through outreach. AIDS Behav 2014 Jul;18(7):1359-1367 [FREE Full text] [doi: 10.1007/s10461-013-0657-z] [Medline: 24287786]
9. Grant JM, Mottet LA, Tanis J, Harrison J, Herman JL, Keisling M. National LGBTQ Task Force. Washington, DC: National Center for Transgender Equality and National Gay and Lesbian Task Force; 2011. Injustice at Every Turn: A Report of the National Transgender Discrimination Survey URL: https://www.thetaskforce.org/wp-content/uploads/2019/07/ntds_full.pdf [accessed 2019-08-12]

10. Centers for Disease Control and Prevention. 2014. CDC-Funded HIV Testing: United States, Puerto Rico, and US Virgin Islands, 2014 URL: https://www.cdc.gov/hiv/pdf/library/reports/cdc-hiv-funded-testing-us-puerto-rico-2014.pdf [accessed 2019-08-12]
11. Centers for Disease Control and Prevention. 2019. Compendium of Evidence-Based Interventions and Best Practices for HIV Prevention URL: https://www.cdc.gov/hiv/research/interventionresearch/compendium/rr/index.html [accessed 2019-08-12]
12. Centers for Disease Control and Prevention. 2019. Effective Interventions: HIV Prevention That Works URL: https://www.cdc.gov/hiv/effective-interventions/index.html [accessed 2019-08-12]
13. Reisner SL, Menino D, Leung K, Gamarel KE. 'Unspoken Agreements': Perceived Acceptability of Couples HIV Testing and Counseling (CHTC) among cisgender men with transgender women partners. AIDS Behav 2019 Feb;23(2):366-374. [doi: 10.1007/s10461-018-2198-y] [Medline: 29936604]
14. Hines DD, Ryan M. It's not just about condoms and sex: using syndemic theory to examine social risks of HIV among transgender women. In: Wright ER, Carnes N, editors. Understanding the HIV/AIDS Epidemic in the United States. Social Disparities in Health and Health Care. Cham, Switzerland: Springer; 2016:99-130.
15. Andrasik MP, Yoon R, Mooney J, Broder G, Bolton M, Votto T, HVTN 505 study team, NIAID HIV Vaccine Trials Network. Exploring barriers and facilitators to participation of male-to-female transgender persons in preventive HIV vaccine clinical trials. Prev Sci 2014 Jun;15(3):268-276 [FREE Full text] [doi: 10.1007/s11121-013-0371-0] [Medline: 23446435]
16. Kosenko K, Rintamaki L, Raney S, Maness K. Transgender patient perceptions of stigma in health care contexts. Med Care 2013 Sep;51(9):819-822. [doi: 10.1097/MLR.0b013e31829fa90d] [Medline: 23929399]
17. Benotsch EG, Zimmerman RS, Cathers L, Heck T, McNulty S, Pierce J, et al. Use of the internet to meet sexual partners, sexual risk behavior, and mental health in transgender adults. Arch Sex Behav 2016 Apr;45(3):597-605. [doi: 10.1007/s10508-014-0432-x] [Medline: 25428577]
18. Krueger EA, Young SD. Twitter: a novel tool for studying the health and social needs of transgender communities. JMIR Ment Health 2015;2(2):pii: e16 [FREE Full text] [doi: 10.2196/mental.4113] [Medline: 26082941]
19. Young SD, Jaganath D. Online social networking for HIV education and prevention: a mixed-methods analysis. Sex Transm Dis 2013 Feb;40(2):162-167 [FREE Full text] [doi: 10.1097/OLQ.0b013e318278bd12] [Medline: 23324979]
20. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
21. Rhodes SD, McCoy TP, Tanner AE, Stowers J, Bachmann LH, Nguyen AL, et al. Using social media to increase HIV testing among gay and bisexual men, other men who have sex with men, and transgender persons: outcomes from a randomized community trial. Clin Infect Dis 2016 Jun 1;62(11):1450-1453 [FREE Full text] [doi: 10.1093/cid/ciw127] [Medline: 26980878]
22. Lightfoot M, Comulada WS, Stover G. Computerized HIV preventive intervention for adolescents: indications of efficacy. Am J Public Health 2007 Jun;97(6):1027-1030. [doi: 10.2105/AJPH.2005.072652] [Medline: 16670219]
23. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190 [FREE Full text] [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
24. Card JJ, Kuhn T, Solomon J, Benner TA, Wingood GM, DiClemente RJ. Translating an effective group-based HIV prevention program to a program delivered primarily by a computer: methods and outcomes. AIDS Educ Prev 2011 Apr;23(2):159-174. [doi: 10.1521/aeap.2011.23.2.159] [Medline: 21517664]
25. Catalani C, Philbrick W, Fraser H, Mechael P, Israelski DM. mHealth for HIV treatment & prevention: a systematic review of the literature. Open AIDS J 2013;7:17-41 [FREE Full text] [doi: 10.2174/1874613620130812003] [Medline: 24133558]
26. Klein CH, Kuhn T, Huxley D, Kennel J, Withers E, Lomonaco CG. Preliminary findings of a technology-delivered sexual health promotion program for black men who have sex with men: quasi-experimental outcome study. JMIR Public Health Surveill 2017 Oct 24;3(4):e78 [FREE Full text] [doi: 10.2196/publichealth.7933] [Medline: 29066422]
27. Noar SM. Computer technology-based interventions in HIV prevention: state of the evidence and future directions for research. AIDS Care 2011 May;23(5):525-533 [FREE Full text] [doi: 10.1080/09540121.2010.516349] [Medline: 21287420]
28. Noar SM, Willoughby JF. eHealth interventions for HIV prevention. AIDS Care 2012;24(8):945-952 [FREE Full text] [doi: 10.1080/09540121.2012.668167] [Medline: 22519523]
29. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134 [FREE Full text] [doi: 10.2196/jmir.3393] [Medline: 24862459]
30. Swanton R, Allom V, Mullan B. A meta-analysis of the effect of new-media interventions on sexual-health behaviours. Sex Transm Infect 2015 Feb;91(1):14-20. [doi: 10.1136/sextrans-2014-051743] [Medline: 25433051]
31. Conserve DF, Jennings L, Aguiar C, Shin G, Handler L, Maman S. Systematic review of mobile health behavioural interventions to improve uptake of HIV testing for vulnerable and key populations. J Telemed Telecare 2017 Feb;23(2):347-359 [FREE Full text] [doi: 10.1177/1357633X16639186] [Medline: 27056905]

32. Fuchs JD, Stojanovski K, Vittinghoff E, McMahan VM, Hosek SG, Amico KR, et al. A mobile health strategy to support adherence to antiretroviral preexposure prophylaxis. AIDS Patient Care STDS 2018 Mar;32(3):104-111 [FREE Full text] [doi: 10.1089/apc.2017.0255] [Medline: 29565183]
33. Dillingham R, Ingersoll K, Flickinger TE, Waldman AL, Grabowski M, Laurence C, et al. PositiveLinks: a mobile health intervention for retention in HIV care and clinical outcomes with 12-month follow-up. AIDS Patient Care STDS 2018 Jun;32(6):241-250 [FREE Full text] [doi: 10.1089/apc.2017.0303] [Medline: 29851504]
34. Sun CJ, Stowers J, Miller C, Bachmann LH, Rhodes SD. Acceptability and feasibility of using established geosocial and sexual networking mobile applications to promote HIV and STD testing among men who have sex with men. AIDS Behav 2015 Mar;19(3):543-552 [FREE Full text] [doi: 10.1007/s10461-014-0942-5] [Medline: 25381563]
35. Arayasirikul S, Wilson EC, Raymond HF. Examining the effects of transphobic discrimination and race on HIV risk among transwomen in San Francisco. AIDS Behav 2017 Sep;21(9):2628-2633 [FREE Full text] [doi: 10.1007/s10461-017-1728-3] [Medline: 28220311]
36. Garofalo R, Kuhns LM, Reisner SL, Mimiaga MJ. Behavioral interventions to prevent HIV transmission and acquisition for transgender women: a critical review. J Acquir Immune Defic Syndr 2016 Aug 15;72(Suppl 3):S220-S225 [FREE Full text] [doi: 10.1097/QAI.0000000000001084] [Medline: 27429186]
37. Horvath KJ, Iantaffi A, Swinburne-Romine R, Bockting W. A comparison of mental health, substance use, and sexual risk behaviors between rural and non-rural transgender persons. J Homosex 2014;61(8):1117-1130 [FREE Full text] [doi: 10.1080/00918369.2014.872502] [Medline: 24380580]
38. LeGrand S, Muessig KE, Pike EC, Baltierra N, Hightow-Weidman LB. If you build it will they come? Addressing social isolation within a technology-based HIV intervention for young black men who have sex with men. AIDS Care 2014;26(9):1194-1200 [FREE Full text] [doi: 10.1080/09540121.2014.894608] [Medline: 24617609]
39. Sun CJ, Anderson KM, Mayer L, Kuhn T, Klein CH. Findings from formative research to develop a strength-based HIV prevention and sexual health promotion mHealth intervention for transgender women. Transgend Health 2019;4(1):350-358 [FREE Full text] [doi: 10.1089/trgh.2019.0032] [Medline: 32042925]
40. Schensul JJ, LeCompte MD. Essential Ethnographic Methods: A Mixed-Methods Approach. Second Edition. Lanham, MD: AltaMira Press; 2012.
41. LeCompte MD, Schensul JJ. Analysis and Interpretation of Ethnographic Data: A Mixed-Methods Approach. Second Edition. Lanham, MD: AltaMira Press; 2012.
42. Nielsen J. Usability Engineering. San Diego, CA: Academic Press; 1993.
43. Dumas JS, Redish JC. A Practical Guide to Usability Testing. Second edition. Exeter, UK: Intellect; 1999.
44. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
45. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
46. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
47. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
48. Hekler EB, King AC, Banerjee B, Robinson T, Alonso M, Cirimele J, et al. A Case Study of BSUED: Behavioral Science-informed User Experience Design. In: Proceedings of the Conference on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada URL: http://v1.personalinformatics.org/docs/chi2011/hekler.pdf
49. 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]
50. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): a framework and toolkit of strategies for the development of more effective digital interventions to change health behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
51. 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]
52. Pagoto S, Bennett GG. How behavioral science can advance digital health. Transl Behav Med 2013 Sep;3(3):271-276 [FREE Full text] [doi: 10.1007/s13142-013-0234-z] [Medline: 24073178]
53. O'Mara-Eves A, Brunton G, Oliver S, Kavanagh J, Jamal F, Thomas J. The effectiveness of community engagement in public health interventions for disadvantaged groups: a meta-analysis. BMC Public Health 2015 Feb 12;15:129 [FREE Full text] [doi: 10.1186/s12889-015-1352-y] [Medline: 25885588]

54. Cyril S, Smith BJ, Possamai-Inesedy A, Renzaho AM. Exploring the role of community engagement in improving the health of disadvantaged populations: a systematic review. Glob Health Action 2015;8:29842 [FREE Full text] [doi: 10.3402/gha.v8.29842] [Medline: 26689460]
55. Centers for Disease Control and Prevention (US). The Principles of Community Engagement. Second Edition. Bethesda, MD: National Institutes of Health; 2011.
56. Cargo M, Mercer SL. The value and challenges of participatory research: strengthening its practice. Annu Rev Public Health 2008;29:325-350. [doi: 10.1146/annurev.publhealth.29.091307.083824] [Medline: 18173388]
57. Israel BA, Schulz AJ, Parker EA, Becker AB. Review of community-based research: assessing partnership approaches to improve public health. Annu Rev Public Health 1998;19:173-202. [doi: 10.1146/annurev.publhealth.19.1.173] [Medline: 9611617]
58. Dankwa-Mullan I, Rhee KB, Williams K, Sanchez I, Sy FS, Stinson N, et al. The science of eliminating health disparities: summary and analysis of the NIH summit recommendations. Am J Public Health 2010 Apr 1;100(Suppl 1):S12-S18. [doi: 10.2105/AJPH.2010.191619] [Medline: 20147660]
59. Nielsen J. Usability Inspection Methods. In: Proceedings of the Conference Companion on Human Factors in Computing Systems. New York, NY: Association for Computing Machinery; 1994 Presented at: CHI'94; April 24 - 28, 1994; Massachusetts, Boston, USA p. 413-414. [doi: 10.1145/259963.260531]
60. Gahagan J, Colpitts E. Understanding and measuring LGBTQ pathways to health: a scoping review of strengths-based health promotion approaches in LGBTQ health research. J Homosex 2017;64(1):95-121. [doi: 10.1080/00918369.2016.1172893] [Medline: 27043161]
61. Operario D, Nemoto T. HIV in transgender communities: syndemic dynamics and a need for multicomponent interventions. J Acquir Immune Defic Syndr 2010 Dec;55(Suppl 2):S91-S93 [FREE Full text] [doi: 10.1097/QAI.0b013e3181fbc9ec] [Medline: 21406995]
62. Cahill S, Valadéz R, Ibarrola S. Community-based HIV prevention interventions that combat anti-gay stigma for men who have sex with men and for transgender women. J Public Health Policy 2013 Jan;34(1):69-81. [doi: 10.1057/jphp.2012.59] [Medline: 23151921]
63. Reback CJ, Ferlito D, Kisler KA, Fletcher JB. Recruiting, linking, and retaining high-risk transgender women into HIV prevention and care services: an overview of barriers, strategies, and lessons learned. Int J Transgend 2015;16(4):209-221 [FREE Full text] [doi: 10.1080/15532739.2015.1081085] [Medline: 27110227]
64. Brennan J, Kuhns LM, Johnson AK, Belzer M, Wilson EC, Garofalo R, Adolescent Medicine Trials Network for HIV/AIDS Interventions. Syndemic theory and HIV-related risk among young transgender women: the role of multiple, co-occurring health problems and social marginalization. Am J Public Health 2012 Sep;102(9):1751-1757 [FREE Full text] [doi: 10.2105/AJPH.2011.300433] [Medline: 22873480]
65. Mayer KH, Grinsztejn B, El-Sadr WM. Transgender people and HIV prevention: what we know and what we need to know, a call to action. J Acquir Immune Defic Syndr 2016 Aug 15;72(Suppl 3):S207-S209 [FREE Full text] [doi: 10.1097/QAI.0000000000001086] [Medline: 27429184]
66. Hwang W, Salvendy G. Number of people required for usability evaluation. Commun ACM 2010;53(5):130-133. [doi: 10.1145/1735223.1735255]
67. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usability Stud 2009;5(1):34-45 [FREE Full text]


## Abbreviations

**CDC:** Centers for Disease Control and Prevention
**EAP:** expert advisory panel
**eHealth:** electronic health
**LGBTQ:** lesbian, gay, bisexual, transgender, and queer
**mHealth:** mobile health
**PrEP:** pre-exposure prophylaxis


*Edited by G Eysenbach; submitted 14.08.19; peer-reviewed by A Tanner, W Goedel; comments to author 02.10.19; revised version received 22.11.19; accepted 15.12.19; published 12.05.20.*

*Please cite as:*
*Sun CJ, Anderson KM, Kuhn T, Mayer L, Klein CH*
*A Sexual Health Promotion App for Transgender Women (Trans Women Connected): Development and Usability Study*
*JMIR Mhealth Uhealth 2020;8(5):e15888*
*URL: https://mhealth.jmir.org/2020/5/e15888*
*doi:10.2196/15888*
*PMID:32396131*

Original Paper

# Efficacy of the Ascure Smoking Cessation Program: Retrospective Study

Ayaka Kato[1,2,3], MSc; Tomoyuki Tanigawa[1,3], MD, MPH; Kohta Satake[1,3,4], MD, MPH, MBA; Akihiro Nomura[1,5,6], MD, PhD

[1]CureApp Institute, Karuizawa, Japan

[2]Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan

[3]CureApp Inc, Tokyo, Japan

[4]Japanese Red Cross Medical Center, Tokyo, Japan

[5]Innovative Clinical Research Center, Kanazawa University, Kanazawa, Ishikawa, Japan

[6]Department of Cardiovascular Medicine, Kanazawa University Graduate School of Medical Sciences, Kanazawa, Japan

**Corresponding Author:**
Akihiro Nomura, MD, PhD
Innovative Clinical Research Center
Kanazawa University
13-1 Takara-machi, Kanazawa
Kanazawa, Ishikawa, 920-8641
Japan
Phone: 81 76 265 2049
Email: anomura@med.kanazawa-u.ac.jp

## Abstract

**Background:** Smoking cessation helps extend a healthy life span and reduces medical expenses. However, the standard 12-week smoking cessation program in Japan has several notable problems. First, only 30% of participants complete this program. Second, participants may choose not to participate unless they have a strong motivation to quit smoking, such as health problems. Third, the program does not provide enough support during the period between clinical visits and after 12 weeks.

**Objective:** This study examined the efficacy of the 24-week ascure program to address the problems of accessibility and continuous support. The program combines online mentoring, over-the-counter pharmacotherapy, and a smartphone app.

**Methods:** Using a retrospective study design, we investigated data for 177 adult smokers who were enrolled in the ascure smoking cessation program between August 2017 and August 2018. The primary outcomes were continuous abstinence rates (CARs) during weeks 9-12 and weeks 21-24. To confirm smoking status, we performed salivary cotinine testing at weeks 12 and 24. We also evaluated the program adherence rate. Finally, we performed exploratory analysis to determine the factors associated with continuous abstinence at weeks 21-24 to provide insights for assisting with long-term continuous abstinence.

**Results:** The CARs of all participants for weeks 9-12 and weeks 21-24 were 48.6% (95% CI 41.2-56.0) and 47.5% (95% CI 40.0-54.8), respectively. Program adherence rates were relatively high throughout (72% at week 12 and 60% at week 24). In the analysis of the factors related to the CAR at weeks 21-24, the number of entries in the app's digital diary and number of educational videos watched during the first 12 weeks were significant factors.

**Conclusions:** The ascure program achieved favorable CARs, and participants showed high adherence. Proactive usage of the smartphone app may help contribute to smoking cessation success in the long-term.

***(JMIR Mhealth Uhealth 2020;8(5):e17270)*** doi:10.2196/17270

**KEYWORDS**

smoking cessation; nicotine dependence; telecare; telemedicine; mHealth; digital therapeutics; mobile phone; smoking cessation program; online counseling

## Introduction

### Background

Smoking is a preventable cause of many health problems, such as cardiovascular diseases, respiratory diseases, and malignant tumors [1,2]. There are more than 20 million smokers in Japan, and deaths from smoking-related diseases are estimated to exceed 12,900 annually [3]. Smoking cessation can greatly contribute to extending a healthy life span and the reduction of medical expenses [3,4]. In Japan, smoking cessation is mainly encouraged through outpatient services. Following a diagnosis of nicotine dependence via interviews and questionnaires at their first visit, patients are eligible for a 12-week standard smoking cessation treatment program that includes 5 clinic visits and an initial screening interview to check exhaled carbon monoxide [5]. In many cases, respiratory physicians and nurses administer the cessation treatment program, which includes the following elements: evaluating the severity of nicotine dependence, measuring the exhaled carbon monoxide concentration, prescribing the appropriate medicine (mostly varenicline or a nicotine patch), and a consultation for quitting smoking, with advice based on behavioral therapy.

This standard smoking cessation program, however, has 3 major problems. First, only about 30% of participants who start the program complete it, often because of the burden of the clinical visits [5]. Second, there is a gap between the number of people who want to quit smoking and those who go to a hospital. The estimated ratio of smokers who participated in the program is only approximately <5% [5,6], although 29.8% of smokers are willing to quit smoking [7]. Participants who choose to join the program might only do so because they have serious health problems or a strong motivation to quit smoking; for example, 38% of outpatient program participants had coexisting diseases [5]. According to pooled analysis from 8 cohort studies in Japan [8], to lower the risk of morbidity from smoking-related cancer to the same level as that of a lifetime non-smoker, male smokers needed over 21 years, and female smokers needed 11 years of abstinence. Therefore, early intervention is important, especially encouraging quitting in young people who might not yet have coexisting diseases or who might be motivated to quit. Third, the continuous abstinence rate (CAR) drops significantly after completing the 12-week program and pharmacological therapy [9]. Given that the decline in CAR plateaus at around week 24, finding a way to assist patients in continuing abstinence beyond 12 weeks could help ensure long-term smoking cessation success [9].

As one way to address the shortcomings of the 12-week program, the “ascure” smoking cessation program was created by CureApp Inc. Its features address some of the concerns related to the standard 12-week program: It is conducted remotely; uses a smartphone app, which is familiar to young people, to provide follow-up assistance; and supports participants for 24 weeks. It also provides professional online mentoring and offers nicotine replacement therapy (over-the-counter medical patches) delivered to the patient’s home. The ascure smartphone app was developed based on research with the CureApp Smoking Cessation app [10-12]. App users have access to tailored guidance in a timely fashion through daily use. The ascure program also offers 6-8 online counseling sessions conducted with experienced nurses and pharmacists.

Smoking cessation programs based on smartphone apps are rapidly increasing globally [13,14]. Despite their popularity, there is limited research on which aspect of the apps predicts smoking cessation. A study on an app employing a behavior change model called acceptance and commitment therapy (ACT) and investigating the associations between a user's smoking cessation and their usage of the app features reported that 2 ACT-specific practices and viewing the cessation plan were the strongest predictors of smoking cessation [15]. Another study analyzed a smoking cessation app designed to deliver the essential features of the United States Clinical Practice Guideline [16]. The authors reported that the number of weeks of active app use (>0 interactions/week) was an important predictor of successful smoking cessation among those using the app. However, no systematic analysis has been done that considers the predictive factors of smoking cessation, especially of app features, when using a combination of online mentoring, over-the-counter medication, and a smartphone app among Japanese smokers.

### Objectives

This study retrospectively examined the clinical efficacy of the ascure smoking cessation program and its impact on continuous smoking cessation. We also used exploratory analysis to determine the factors associated with continuous abstinence at weeks 21-24 to gain insights for supporting long-term continuous abstinence.

## Methods

### Study Design

This was a retrospective study evaluating the efficacy of the ascure smoking cessation program. We conducted this study in compliance with the Declaration of Helsinki, and all other applicable laws and guidelines in Japan. All study procedures were reviewed and approved by the Kanazawa University Institutional Review Board (2019-023 (3058)).

### Participants

We assessed 177 adult smokers participating in the ascure smoking cessation program in Japan from August 2017 to August 2018. All participants had their own iOS or Android smartphones, desired to quit smoking, and were members of one of 16 corporate health insurance societies. Participants were recruited mostly via self-selection. Some became aware of this program through individual recommendations from professionals (public health nurses) at medical checkups. Advertisements by email, post, or leaflets distributed at hospitals were used for recruitment. We included participants who smoked >10 cigarettes per day and who agreed to the study’s privacy policy by providing written informed consent for data to be used in our analysis. We excluded participants who had difficulty using their smartphones according to the instructions or who were diagnosed with a mental illness.

## Outcomes

The primary outcomes of this study were participants' CARs during weeks 9-12 and weeks 21-24. During the online video sessions at weeks 12 and 24, mentors asked participants whether they had smoked during the month prior to the current session and asked them to perform salivary cotinine testing using iScreen (Cotinine Oral Fluid Screening Device, Abbott Diagnostics Medical Co, Ltd, Tokyo, Japan). The results of the test were confirmed visually via video. This allowed supplementation of the self-reported continuous abstinence via salivary cotinine testing. We also assessed adherence to the program based on how many weeks participants took part in the online sessions (weeks 1, 2, 4, 8, 12, and 24). Furthermore, we used exploratory analysis to determine the variable factors associated with smoking cessation success at week 24 (demographic data, smoking history, and app usage) to provide potential insights for maintaining long-term continuous abstinence.

## Ascure Smoking Cessation App

The ascure smoking cessation program is a 24-week completely remote, online program including 6 online sessions with the exclusive ascure smoking cession smartphone app. The app was developed by CureApp Inc (Tokyo, Japan). As noted earlier, the app's contents were developed based on the findings of clinical studies using the CureApp Smoking Cessation app for patients diagnosed with nicotine dependence (Multimedia Appendix 1). The ascure app is compatible with both iOS and Android smartphones and meets the software inspection criteria and security requirements of the Apple App Store and Google Play. CureApp Inc provided participants with the app activation codes to begin use. The participants downloaded the app, activated the app using the codes provided, and entered their personal demographic information including age, gender, years of smoking, and estimated number of cigarettes smoked per day. Their information was securely stored on a cloud system and utilized to support the smoking cessation of each participant in their personalized counseling.

The app (see Figure 1) comprises 4 steps to maximize the therapeutic effect of the medications and online mentoring sessions for smoking cessation: learning, exercise, keeping a record, and rescue.

**Figure 1.** Scheme of the ascure smoking cessation program.


### *Learning: Educational Video Tutorials to Help Users Quit Smoking*

Participants were encouraged to watch 1-3–minute video tutorials on nicotine dependence that provided useful tips on smoking cessation based on behavioral therapy. A total of 24 video tutorials were provided during the 24-week smoking cessation program.

### *Exercise: A Personalized To-Do List to Change Habits*

After being exposed to the behavioral therapy through the tutorials, participants could make a personalized to-do list for quitting smoking. For example, they could list alternative behaviors such as brushing their teeth or stretching when they craved a cigarette. When participants carried out the registered alternative behavior, they could record it in the digital diary provided in the app.

### *Keeping a Record: Digital Diary of Smoking Cessation*

The app instructed participants to keep a digital diary on their smoking cessation status, physical condition, medication use, and any adverse events, either by selecting from prepopulated options or by writing entries in free form.

### *Rescue: Interactive Chat Sessions*

Whenever participants experienced cravings or withdrawal symptoms, they could tap the "call" button to send a message to a personalized chatbot. The chatbot immediately replied and provided personalized advice on how to deal with the symptoms. The chatbot also provided encouraging messages on smoking

cessation to app users to remind the participants of their motivation to quit smoking.

### Online Mentoring

Throughout the 24-week program, 6-8 interactive online sessions with experienced mentors (normally professional nurses or pharmacists) were provided. Participants met with mentors via a web-based counseling system at each planned visit (weeks 1, 2, 4, 8, and 12, as in the standard smoking cessation outpatient program, with an additional session in week 24). During the online sessions, the mentors asked participants about their smoking status and any difficulties they encountered. Participants were also provided with guidance on overcoming psychological dependence. Each online mentor in the ascure program had more than 3 years of experience as a nurse, public health nurse, dietitian, or pharmacist. All instructors joined the Japan Society of Tobacco Control or The Japanese Association of Smoking Control Science and were qualified by either society before their initial mentoring session. In-house training was conducted for 1 month to learn the system for smoking cessation.

### Medication

As part of the program, participants received over-the-counter nicotine patches or nicotine gum as a smoking cessation aid, which was delivered via parcel from a pharmacy (ascure store [former Nihonbashi Smart Clinic], Tokyo, Japan) to their homes. This nicotine replacement treatment was typically used until week 8.

### Data Collection

We collected basic information on each participant via the app, including their age, gender, years of smoking, number of cigarettes smoked per day, and screening of tobacco/nicotine dependence using methods such as the Tobacco Dependence Screener (TDS) score [17], Fagerström Test for Nicotine Dependence score [18], and Kano Test for Social Nicotine Dependence (KTSND) score [19], the metrics of which focus on their psychological dependence. We also gathered app usage statistics, such as the number of days that participants updated their diaries, number of "call" button presses, and number of educational videos viewed.

### Statistical Analysis

Continuous variables are presented as mean (SD) or median (interquartile range [IQR]), depending on their distribution. Categorical variables are presented as the number (percentage). For the exploratory factor analysis of the predictors of CAR at week 24, we conducted multivariable logistic regression analysis, selecting from among the standardized variables (demographic data, smoking history, and app usage evaluated between weeks 0 and 12), while adjusting for age, gender, and TDS score. In multiple testing, the Bonferroni method was used to adjust the *P* values. *P* values $<3.0 \times 10^{-3}$ (0.05 divided by 15) were considered to be significant. There were 36 missing values for the TDS score and 40 missing values each for the Fagerström Test for Nicotine Dependence and KTSND. To address these missing data in the explanatory variables, we used a multiple imputation method. We used the Pearson Chi-square test for subpopulation analysis. All calculations and analyses were performed using *R* software version 3.5.2 (*R* Foundation for Statistical Computing, Vienna, Austria); packages used were 'data.table,' 'stats,' and 'mice'.

## Results

### Baseline Characteristics

Table 1 shows the baseline characteristics of the 177 participants.

**Table 1.** Baseline characteristics of the 177 participants.

| Characteristics | Values | Range |
|---|---|---|
| Age (years), mean (SD) | 44.6 (9.7) | 26-63 |
| Female gender, n (%) | 65 (36.7) | N/A |
| Brinkman index, mean (SD) | 362 (228) | 20-1260 |
| Cigarettes per day, mean (SD) | 16.2 (6.4) | 2-40 |
| Years of smoking, mean (SD) | 22.0 (9.8) | 2-42 |
| Number of smoking cessation attempts before the trial, median (IQR[a]) | 1 (3) | 0-11 |
| TDS[b] score, median (IQR) | 8 (3) | 0-10 |

[a]IQR: interquartile range.

[a]TDS: Tobacco Dependence Screener.

### Efficacy of the Ascure Program

Program adherence was 71.8% (127/177) at week 12 and 59.9% (106/177) at week 24. Within the first 4 weeks of the program, 23 people did not book the subsequent online mentoring sessions, and 11 and 16 people, respectively, lost contact with their mentors during the week 4 and week 8 sessions. The biochemically validated CAR during weeks 9-12 was 48.6% (86/177; 95% CI 41.2-56.0; Table 2). At weeks 9-12, 86 people had successfully quit smoking, while 50 people could not be evaluated because they dropped out of the program. The biochemically validated CAR during weeks 21-24 was 47.4% (84/177; 95% CI 40.0-54.8). We confirmed continuous adherence for 84 of the 106 individuals who completed the full 24-week program. Of these, 75 people also succeeded in 1 month of smoking cessation during weeks 9-12, and 9 people who had been counted as unsuccessful at week 12 became successful at week 24. No participants reported adverse events,

including app-related events, during the program other than skin rashes and nausea caused by nicotine patches.

**Table 2.** Efficacy of the ascure program.

| Measurement | Efficacy rate, n (%) | 95% CI |
|---|---|---|
| CAR[a] at weeks 9-12 | 86 (48.6) | 41.2-56.0 |
| CAR at weeks 21-24 | 84 (47.4) | 40.0-54.8 |

[a]CAR: continuous abstinence rate.

## Predictors of CAR at Weeks 21-24

Next, we evaluated the predictors of CAR during weeks 21-24 (Table 3). The candidate factors were the demographic variables, smoking history, and app usage. The learning comportment of the app was quantified by the number of educational video tutorials watched. A personalized to-do list to change habits included the number of times environmental triggers were avoided, number of behavioral patterns changed, number of alternative behaviors, and number of instances of assertiveness training. The number of diary entries and reports of poor physical condition reflected the use of the recording function of this app. The use of the interactive chat was measured by the number of "like" button presses and nurse "call" button presses. The multivariable analyses revealed that the number of days participants wrote in their digital diaries and number of educational video tutorials watched during the first 12 weeks were significantly associated with CAR during weeks 21-24.

Moreover, we calculated the CARs at weeks 21-24 separately by subgroups of the number of days that participants wrote in their diaries. Participants who wrote diary entries more than the median number (43 times; range, 0-77 times) during the first half of the program achieved higher CARs (59/88, 67%) than those who did not (25/89, 28%; $\chi^1_{177} = 25.39$, $P<.001$) when counting dropout as a failure. Even within the group of 106 people who completed all 24 weeks, the subgroup who wrote more diary entries than the median number (62.5 times) had a higher success rate (47/53, 88.7%) compared with those who did not (37/53, 69.8%; $\chi^1_{106} = 4.65$, $P=.03$).

Participants who watched educational videos more than the median number (6 times; range, 0-24 times) also had significantly higher CARs (55/86, 64.0% vs. 29/91, 31.8%) when counting dropout as a failure ($\chi^1_{177} = 16.99$, $P<.001$). However, there was no significant difference between success or failure among people who completed all 24 weeks ($\chi^1_{177} = 2.19$, $P=.14$).

**Table 3.** Results of multivariable logistic regression analysis to determine potential predictors of smoking cessation success at week 24.

| Variable | Odds ratio | 95% CI | *P* value | AIC[a] |
|---|---|---|---|---|
| **Demographic data** | | | | |
| Age | 1.577 | 1.159-2.174 | .004 | 239 |
| Sex (male: 0, female: 1) | 0.552 | 0.289-1.039 | .068 | 239 |
| **Smoking history[b]** | | | | |
| Years of smoking | 1.405 | 0.840-2.409 | .197 | 239 |
| Number of cigarettes/day | 1.011 | 0.743-1.375 | .944 | 241 |
| Number of smoking cessation attempts before the trial | 0.838 | 0.604-1.140 | .268 | 240 |
| TDS[c] | 0.752 | 0.537-1.036 | .087 | 238 |
| FTND[d] | 0.791 | 0.537-1.081 | .146 | 239 |
| KTSND[e] | 0.747 | 0.523-1.043 | .095 | 238 |
| **Application usage[f]** | | | | |
| Educational video tutorials watched | 2.178 | 1.550-3.126 | <.001 | 219 |
| Times environmental triggers were avoided | 2.441 | 1.367-5.394 | .011 | 229 |
| Times behavioral patterns were changed | 3.184 | 1.550-9.029 | .009 | 226 |
| Alternative behaviors | 2.514 | 1.404-5.922 | .010 | 228 |
| Instances of assertiveness training | 1.560 | 1.017-2.896 | .089 | 236 |
| Diary entries | 2.661 | 1.866-3.894 | <.001 | 208 |
| Reports of poor physical condition | 1.172 | 0.860-1.630 | .320 | 239 |
| “Like” button presses | 1.638 | 1.108-2.673 | .026 | 237 |
| Nurse call button presses | 1.510 | 1.059-2.366 | .040 | 235 |

[a]AIC: Akaike information criterion.

[b]Adjusted by age and gender.

[c]TDS: Tobacco Dependence Screener.

[d]FTND: Fagerström Test for Nicotine Dependence.

[e]KTSND: Kano Test for Social Nicotine Dependence.

[f]Adjusted by age, gender, and TDS.

## Exploratory Analyses

As a reference, we reviewed the efficacy of the standard smoking cessation program. There are several differences in the baseline characteristics between the ascure program and the standard program. Compared with the national survey (Table 4), the ascure program participants were relatively young, more often female, and had low dependence, as represented by the Brinkman index.

While only 29.8% (390/1308) of smoking cessation outpatients completed the standard 12-week treatment, ascure participants had a higher program adherence rate of 71.8% (127/177) at week 12 and 59.9% (106/177) even at week 24.

Although the efficacy of the 2 programs cannot be compared directly, both programs measured efficacy with a 1-month CAR at weeks 9-12. In the case of the outpatient service, the self-reported 1-month continuous abstinence was confirmed by exhaled carbon monoxide concentration (<8 ppm) at the fifth clinical visit.

Among the participants continuing the ascure program until week 12, the CAR at week 12 was 67.7% (86/127; 95% CI 59.4-76.0), whereas the CAR among individuals who completed the 12-week outpatient program was 82.1% (320/390) [5]. However, in contrast to the 12-week outpatient program, the ascure program assisted participants for up to 24 weeks through the smartphone app. The CAR at weeks 21-24 among people who completed the ascure program was 79.2% (84/106; 95% CI 71.4-87.1). Importantly, when counting dropouts as failures, the CAR of all ascure participants at weeks 9-12 (86/177, 48.6%) was twice that of outpatients in the standardized smoking cessation program (320/1308, 24.5%) [5].

**Table 4.** Baseline characteristics of the participants.

| Characteristic | Standard program, n=1308 | ascure program, n=177 |
|---|---|---|
| Age (years), mean (SD) | 49.0 (14.5) | 44.6 (9.7) |
| Female sex, n (%) | 400 (30.6) | 65 (36.7) |
| Brinkman index, mean (SD) | 634.1 (448.3) | 362 (228) |
| Cigarettes per day, mean (SD) | 22.8 (10.2) | 16.2 (6.4) |
| Years of smoking, mean (SD) | 27.5 (13.4) | 22.0 (9.8) |
| TDS[a] score, median (IQR[b]) | 8 (N/A[c]) | 8 (3) |

[a]TDS, Tobacco Dependence Screener.

[b]IQR: interquartile range.

[c]IQR not reported in the publication.

# Discussion

## Principal Findings

This study evaluated the efficacy of the ascure smoking cessation program, which is characterized by online mentoring, a smartphone app, over-the-counter medications, and an extended follow-up period (24 weeks). First, we found that the CAR was 48.6% during weeks 9-12 and 47.5% during weeks 21-24. We also observed high adherence rates: 72% at week 12 and 60% at week 24. Unlike previous reports [9], CARs did not decrease significantly between weeks 12 and 24. Second, using multivariable logistic regression analysis, we found that the number of days of digital diary entries and number of educational videos watched during the first 12 weeks were significantly associated with the CAR at weeks 21-24. The number of days of digital diary entries was significantly different between people who succeeded or failed in 1-month continuous abstinence at week 24 even when dropouts were excluded.

## Comparison With Prior Work

This study has 3 important findings. First, the ascure smoking cessation program achieved moderately favorable CARs in comparison with the current standard Japanese outpatient program, mainly due to the adherence rate. While only about 30% of participants completed the outpatient program [5], the completely remote ascure program saw a higher adherence rate, even at week 24. Participants in the ascure program achieved reasonable CARs without face-to-face advice from medical doctors; without taking varenicline, a prescription drug that blocks the pleasant effects of nicotine; and without the need to visit smoking cessation clinics. This program is intended to provide an alternative to the standard program for those who cannot access the standard program, including busy professionals. This program could be convenient for people who do not visit hospitals frequently, who do use a smartphone frequently, and who are familiar with web-conferencing. On the other hand, the outpatient service is a better solution for those who have other serious health problems or have experienced withdrawal symptoms. The differentiating aspects of ascure from other app-based smoking cessation programs in the world include the contents of the app, professional mentoring, and pharmacotherapy. Bricker et al [20] evaluated a smoking cessation app delivering ACT in 99 adult smokers and found that 11% had achieved 30-day point prevalence abstinence in an intent-to-treat (ITT) analysis at the 2-month follow-up. The app assessed by Iacoviello et al [16] has a series of missions and personalized messages that adhere to the United States Clinical Practice Guideline, and 26.2% of the ITT sample (109/416) reported 30-day abstinence from smoking after 8 weeks. Another study evaluated a comprehensive, multiphase digital smoking cessation program that includes a mobile carbon monoxide breath sensor and text-based human coaching and reported a 30-day point prevalence abstinence of 27.6% (88/319, ITT) [21]. The outcomes could have been strongly affected by the difference in the country the study was conducted in and ethnicity. Also, the study design, program duration, baseline characteristics, and outcome measurement are not consistent among these programs, including the ascure program. We acknowledge that the huge differences limit the comparability of these programs, but the efficacy of the ascure program is favorable when assessing 1-month smoking cessation success.

Second, proactive use of the smartphone app (eg, keeping a digital diary and watching educational videos) during the first 12 weeks of the program was significantly associated with CARs during weeks 21-24. Keeping a smoking cessation diary is recommended by various Japanese academic societies (eg, the Cardiovascular Society, Lung Cancer Society, Cancer Society, and Respiratory Society) [22]. This study further supports the concept that writing diaries is associated with an increased success rate in smoking cessation. The number of diary entries might reflect an individual's willingness to quit smoking and therefore their commitment to the program. A study comparing two other smartphone-based smoking cessation interventions — SmartQuit and Quit guide — reported that the number of times the application was opened was a significant predictor of smoking cessation [23]. In addition, another study on the app-based program, Clickotine, also found that the number of weeks with more than 1 interaction in the app was associated with smoking cessation success [16]. Participant engagement has been found to be a consistent predictor of smoking cessation in several programs, and interactivity might be a key factor in improving engagement [15,16,23-26]. The proactive use of a smartphone app as an indicator of patient motivation might be a predictive factor of better outcomes of smoking cessation programs.

Third, the ascure program seemed to attract more relatively young, female participants with low dependence, as represented by the Brinkman index. This may imply that, because of the completely remote aspect of the ascure program, it could recruit people who do not utilize the outpatient service voluntarily.

### Strengths and Limitations

This study was the first to examine the efficacy of a novel smoking cessation program for members of a Japanese health insurance society. We used a biochemically validated smoking status test to assess the efficacy of the program. However, this study has several limitations. First, the absence of a control group is a significant problem when assessing the efficacy of the program. Our finding that a predictor of smoking cessation was the intensive use of the program could merely reflect that motivated patients made more intensive use of the program. Future studies, such as crossover trials with a control group, are necessary to better evaluate the efficacy of the program. Second, several factors could limit the generalizability of the findings: participants' level of motivation to engage in the program, participants' sociodemographic status (all were members of corporate health insurance societies), and access to technology (ie, smartphones). Third, we evaluated 1-month continuous abstinence (as does the outpatient service in Japan). This limits the comparability with internet-based programs in other countries. Fourth, data on adherence to nicotine patches and the quality of video counseling, which might affect outcomes, were not collected. Finally, it may be necessary to evaluate smoking status even further into the future, such as at week 52 or beyond, to confirm the long-term efficacy of this program.

### Conclusion

The completely remote ascure smoking cessation program achieved favorable smoking cessation success rates while improving accessibility and adherence to a cessation program. Moreover, proactive use of the smartphone app may have contributed to successful smoking cessation over a relatively long period. The combination of internet-based counseling, a smartphone app, and over-the-counter medications might be a viable solution for long-term smoking cessation.

## Acknowledgments

We must thank all staff and participants involved in this program. We would like to thank Editage (www.editage.com) for English language editing. This study was funded by CureApp Inc and the Masason Foundation. The funders of the study had no role in study design, data analysis, data interpretation, or writing of the report. CureApp Inc supported data collection.

## Authors' Contributions

AK and AN conceived the study. AK performed the data analysis, which was reviewed and interpreted by AN, TT, and KS. AK and AN wrote the manuscript, and TT and KS reviewed and commented on the contents. All authors contributed to the preparation of the article and approved the final version. AK and AN had full access to all the data in the study. AN had final responsibility for the decision to submit for publication.

## Conflicts of Interest

AN received consulting fees from CureApp, Inc, Japan. AK and TT are employees of CureApp, Inc, Japan. KS is the founder and a shareholder of CureApp, Inc.

Multimedia Appendix 1
Interface of the ascure smoking cessation app.
[MP4 File (MP4 Video), 21106 KB - mhealth_v8i5e17270_app1.mp4 ]

## References

1. Ikeda N, Inoue M, Iso H, Ikeda S, Satoh T, Noda M, et al. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. In: PLoS Med. Adult mortality attributable to preventable risk factors for non-communicable diseases and injuries in Japan: a comparative risk assessment. PLoS Med; Jan 2012:e1001160.
2. World Health Organization. WHO Report on the Global Tobacco Epidemic, 2017 URL: http://apps.who.int/iris/bitstream/handle/10665/255874/9789241512824-eng.pdf [accessed 2019-04-19]
3. Ikeda N, Saito E, Kondo N, Inoue M, Ikeda S, Satoh T, et al. What has made the population of Japan healthy? The Lancet 2011 Sep;378(9796):1094-1105. [doi: 10.1016/s0140-6736(11)61055-6]
4. Igarashi A, Goto R, Suwa K, Yoshikawa R, Ward AJ, Moller J. Cost-Effectiveness Analysis of Smoking Cessation Interventions in Japan Using a Discrete-Event Simulation. Appl Health Econ Health Policy 2016 Feb 23;14(1):77-87 [FREE Full text] [doi: 10.1007/s40258-015-0204-3] [Medline: 26597111]
5. Ministry of Health, Labour and Welfare, Health Service Bureau General Affairs Division. Report on the Efficacy of Smoking Cessation Treatment Based on Nicotine Dependency Management Fee URL: https://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000192293.pdf [accessed 2020-04-27]

6. Japan Tobacco Inc National survey of the ratio of tobacco smokers. URL: https://www.jti.co.jp/investors/library/press_releases/2018/0730_01.html [accessed 2020-02-24]
7. Ministry of Health, Labour and Welfare, National Health and Nutrition Survey. URL: https://www.mhlw.go.jp/content/000451758.pdf [accessed 2020-02-24]
8. Saito E, Inoue M, Tsugane S, Ito H, Matsuo K, Wakai K, Research Group for the Development Evaluation of Cancer Prevention Strategies in Japan. Smoking cessation and subsequent risk of cancer: A pooled analysis of eight population-based cohort studies in Japan. Cancer Epidemiol 2017 Dec;51:98-108. [doi: 10.1016/j.canep.2017.10.013] [Medline: 29102692]
9. Nakamura M, Oshima A, Fujimoto Y, Maruyama N, Ishibashi T, Reeves KR. Efficacy and tolerability of varenicline, an alpha4beta2 nicotinic acetylcholine receptor partial agonist, in a 12-week, randomized, placebo-controlled, dose-response study with 40-week follow-up for smoking cessation in Japanese smokers. Clin Ther 2007 Jun;29(6):1040-1056. [doi: 10.1016/j.clinthera.2007.06.012] [Medline: 17692720]
10. Masaki K, Tateno H, Kameyama N, Morino E, Watanabe R, Sekine K, et al. Impact of a Novel Smartphone App (CureApp Smoking Cessation) on Nicotine Dependence: Prospective Single-Arm Interventional Pilot Study. JMIR Mhealth Uhealth 2019 Feb 19;7(2):e12694 [FREE Full text] [doi: 10.2196/12694] [Medline: 30777848]
11. Nomura A, Tanigawa T, Muto T, Oga T, Fukushima Y, Kiyosue A, et al. Clinical Efficacy of Telemedicine Compared to Face-to-Face Clinic Visits for Smoking Cessation: Multicenter Open-Label Randomized Controlled Noninferiority Trial. J Med Internet Res 2019 Apr 26;21(4):e13520 [FREE Full text] [doi: 10.2196/13520] [Medline: 30982776]
12. Masaki K, Tateno H, Nomura A, Muto T, Suzuki S, Satake K, et al. A randomized controlled trial of a smoking cessation smartphone application with a carbon monoxide checker. NPJ Digit Med 2020;3(1):35 [FREE Full text] [doi: 10.1038/s41746-020-0243-5] [Medline: 32195370]
13. Abroms LC, Lee Westmaas J, Bontemps-Jones J, Ramani R, Mellerson J. A content analysis of popular smartphone apps for smoking cessation. Am J Prev Med 2013 Dec;45(6):732-736 [FREE Full text] [doi: 10.1016/j.amepre.2013.07.008] [Medline: 24237915]
14. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun 19;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
15. Heffner JL, Vilardaga R, Mercer LD, Kientz JA, Bricker JB. Feature-level analysis of a novel smartphone application for smoking cessation. Am J Drug Alcohol Abuse 2015 Jan 14;41(1):68-73 [FREE Full text] [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
16. Iacoviello BM, Steinerman JR, Klein DB, Silver TL, Berger AG, Luo SX, et al. Clickotine, A Personalized Smartphone App for Smoking Cessation: Initial Evaluation. JMIR Mhealth Uhealth 2017 Apr 25;5(4):e56 [FREE Full text] [doi: 10.2196/mhealth.7226] [Medline: 28442453]
17. Kawakami N, Takatsuka N, Inaba S, Shimizu H. Development of a screening questionnaire for tobacco/nicotine dependence according to Icd-10, Dsm-III-r, And Dsm-IV. Addictive Behaviors 1999 Mar;24(2):155-166. [doi: 10.1016/s0306-4603(98)00127-0]
18. Heatherton T, Kozlowski L, Frecker R, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
19. Otani T, Yoshii C, Kano M, Kitada M, Inagaki K, Kurioka N, et al. Validity and reliability of Kano Test for Social Nicotine Dependence. Ann Epidemiol 2009 Nov;19(11):815-822. [doi: 10.1016/j.annepidem.2009.03.016] [Medline: 19457683]
20. Bricker JB, Copeland W, Mull KE, Zeng EY, Watson NL, Akioka KJ, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2017 Jan 01;170:37-42 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
21. Marler JD, Fujii CA, Utley DS, Tesfamariam LJ, Galanko JA, Patrick H. Initial Assessment of a Comprehensive Digital Smoking Cessation Program That Incorporates a Mobile App, Breath Sensor, and Coaching: Cohort Study. JMIR Mhealth Uhealth 2019 Feb 04;7(2):e12609 [FREE Full text] [doi: 10.2196/12609] [Medline: 30670372]
22. Cardiovascular Society, Lung Cancer Society, Cancer Society, and the Respiratory Society. The standard protocol for smoking cessation treatment, 6th edition URL: http://www.j-circ.or.jp/kinen/anti_smoke_std/pdf/anti_smoke_std_rev6.pdf [accessed 2019-04-19]
23. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
24. Civljak M, Sheikh A, Stead LF, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010 Sep 08(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
25. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]
26. Whittaker R, Merry S, Dorey E, Maddison R. A development and evaluation process for mHealth interventions: examples from New Zealand. J Health Commun 2012 May 02;17 Suppl 1(sup1):11-21. [doi: 10.1080/10810730.2011.649103] [Medline: 22548594]

## Abbreviations

**ACT:** acceptance and commitment therapy
**CAR:** continuous abstinence rate
**FTND:** Fagerström Test for Nicotine Dependence
**IQR:** interquartile range
**ITT:** intention-to-treat
**KTSND:** Kano Test for Social Nicotine Dependence
**TDS:** Tobacco Dependence Screener

*Edited by G Eysenbach; submitted 05.12.19; peer-reviewed by C Mavrot, Y Liao, JF Lisón, F Sartor; comments to author 06.01.20; revised version received 01.03.20; accepted 09.04.20; published 14.05.20.*

*Please cite as:*
*Kato A, Tanigawa T, Satake K, Nomura A*
*Efficacy of the Ascure Smoking Cessation Program: Retrospective Study*
*JMIR Mhealth Uhealth 2020;8(5):e17270*
*URL: https://mhealth.jmir.org/2020/5/e17270*
*doi:10.2196/17270*
*PMID:32406856*

Review

# Use of Apps to Promote Childhood Vaccination: Systematic Review

Caroline de Cock[1], BSc, MSc; Michelle van Velthoven[1], BSc, MSc, PhD; Madison Milne-Ives[1], BAS, MSc; Mary Mooney[2], MA, RM, RNT, RGN, HDip, MSc, PhD; Edward Meinert[1,3], MA, MSc, MBA, MPA, PhD

[1]Digitally Enabled PrevenTative Health Research Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom

[2]School of Nursing & Midwifery, Trinity College Dublin, Dublin, Ireland

[3]Department of Primary Care and Public Health, Imperial College London, London, United Kingdom

**Corresponding Author:**
Edward Meinert, MA, MSc, MBA, MPA, PhD
Digitally Enabled PrevenTative Health Research Group
Department of Paediatrics
University of Oxford
Children's Hospital
John Radcliffe Hospital
Oxford, OX3 9DU
United Kingdom
Phone: 44 7824446808
Email: e.meinert14@imperial.ac.uk

## Abstract

**Background:** Vaccination is a critical step in reducing child mortality; however, vaccination rates have declined in many countries in recent years. This decrease has been associated with an increase in the outbreak of vaccine-preventable diseases. The potential for leveraging mobile platforms to promote vaccination coverage has been investigated in the development of numerous mobile apps. Although many are available for public use, there is little robust evaluation of these apps.

**Objective:** This systematic review aimed to assess the effectiveness of apps supporting childhood vaccinations in improving vaccination uptake, knowledge, and decision making as well as the usability and user perceptions of these apps.

**Methods:** PubMed, Excerpta Medica Database (EMBASE), Web of Science, Cochrane Central Register of Controlled Trials, ClinicalTrials.gov, and Education Resources Information Center (ERIC) databases were systematically searched for studies published between 2008 and 2019 that evaluated childhood vaccination apps. Two authors screened and selected studies according to the inclusion and exclusion criteria. Data were extracted and analyzed, and the studies were assessed for risk of bias.

**Results:** A total of 28 studies evaluating 25 apps met the inclusion criteria and were included in this analysis. Overall, 9 studies assessed vaccination uptake, of which 4 reported significant benefits ($P<.001$ or $P=.03$) of the implementation of the app. Similarly, 4 studies indicated a significant ($P\leq.054$) impact on knowledge and on vaccination decision making. Patient perceptions, usability, and acceptability were generally positive. The quality of the included studies was found to be moderate to poor, with many aspects of the methodology being unclear.

**Conclusions:** There is little evidence to support the use of childhood vaccination apps to improve vaccination uptake, knowledge, or decision making. Further research is required to understand the dichotomous effects of vaccination-related information provision and the evaluation of these apps in larger, more robust studies. The methodology of studies must be reported more comprehensively to accurately assess the effectiveness of childhood vaccination apps and the risk of bias of studies.

**International Registered Report Identifier (IRRID):** RR2-10.2196/16929

***(JMIR Mhealth Uhealth 2020;8(5):e17371)*** doi:10.2196/17371

**KEYWORDS**

vaccination; vaccination coverage; mobile apps; infant; childhood vaccination; immunization; smartphone technology; mobile phone

## Introduction

### Background

In 2018, it was estimated that immunization prevented 2 to 3 million deaths each year, yet over 19 million children worldwide under the age of 1 year did not receive basic vaccines [1]. Most of these children lived in developing countries, where access to vaccines and antenatal services is somewhat limited [1]. Nevertheless, an increase in vaccine-preventable disease outbreaks has also been identified in developed countries, which is associated with declining vaccination uptake [2-4]. Immunization coverage of 9 routine childhood vaccinations declined in England by 0.2% to 1% during 2018 to 2019, compared with the previous year, and 1.3% of children born in 2015 in the United States received no vaccinations by the age of 2 years, compared with 0.9% of those born in 2011 [4,5]. A number of studies have investigated the reasons for vaccine refusal among parents and caregivers and have revealed that religious, philosophical, and personal beliefs, coupled with safety concerns and insufficient information, were the most commonly cited reasons [6]. The now-refuted evidence linking measles-mumps-rubella with autism was seen to cause a 2% decrease in the uptake of measles-mumps-rubella vaccinations [7]. Furthermore, the widespread adoption of the human papillomavirus (HPV) vaccine has been thwarted by religious and cultural barriers [8,9].

Despite the low mortality rate of vaccine-preventable diseases, various sociodemographic groups, including young children and elderly or immunocompromised individuals, are at risk of serious, sometimes fatal, complications [10]. The outbreaks of vaccine-preventable diseases can be minimized by maintaining herd immunity, which varies from 75% to 97% vaccination coverage, depending on the disease and setting in question [11]. As such, it is crucial that there is adequate provision of correct and comprehensive information, resources, and reminders to encourage parents and caregivers to obtain complete and timely vaccinations for their children. Many informational, behavioral, and environmental initiatives in various settings have been implemented to improve the uptake of childhood vaccinations [12-14]; however, the scalability and sustainability of these programs have been problematic [12].

With the increasing utilization and accessibility of mobile devices, digital technologies have shown promise in effectively disseminating information to diverse and diffuse populations and rolling out community-wide initiatives [10,15]. A World Health Organization survey on mobile health (mHealth) revealed that 83% of member states offered at least one service in 2011 [16]. mHealth has been investigated by multiple private and public organizations to support the uptake of vaccinations, including vaccination information websites and mobile apps, hereafter referred to as apps. These apps have various functions designed to support health care providers, caregivers, and, in some cases, children to access vaccine-related information, recommended immunization schedules, store vaccination records, and receive appointment reminders. There are now over 200 vaccination-related apps available on the App Store [17]. A systematic review conducted in 2015 discussing the design of vaccination reminder apps reviewed 2 studies on mobile reminder apps [18]. However, a comprehensive review of the effectiveness and usability of childhood vaccination-related apps is yet to be conducted.

### Objectives

This study aimed to systematically review the evidence on the use of apps to support childhood vaccination uptake, information storage, and record sharing as well as to investigate the usability and user perceptions of these apps.

## Methods

This systematic review was conducted following, where possible, the Cochrane collaboration [19] and the Centre for Review and Dissemination [20] methodologies for conducting systematic reviews.

### Database Search

Full methods for this review have been published in detail in a systematic review protocol [21]. This systematic review was registered with the International Prospective Register of Systematic Reviews (CRD42019156583). The Participant, Intervention, Comparison, and Outcome framework was used to develop the search strategy [22], which was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols [23]. The search strategy was tweaked to ensure that a defined set of known references was returned without retrieving an unmanageably large number of studies. This primarily involved the selection and amendment of wildcard terms to ensure that irrelevant terms were not included; for example, *immun** retrieved papers related to immunology; hence, this was amended to *immuni** to make this more specific to immunizations. No study design filter was used as both quantitative and qualitative studies were included. The search strategy was finalized and tailored to different databases in consultation with a medical librarian. PubMed, Excerpta Medica Database (EMBASE, Ovid), Web of Science, Cochrane Central Register of Controlled Trials (CENTRAL), ClinicalTrials.gov, and Education Resources Information Center (ERIC) databases were searched. The search terms were grouped into 3 themes—vaccinations, mobile apps, and children—which were subsequently searched with the following structure: vaccinations (Medical Subject Headings, MeSH OR Keywords) AND mobile applications (MeSH OR Keywords) AND children (MeSH OR Keywords). The full search strategy is shown in Multimedia Appendix 1. The search took place on October 23, 2019.

### Inclusion and Exclusion Criteria

This systematic review aimed to assess apps designed to support childhood vaccination uptake. As such, the search was limited to studies conducted during or after 2008, when the first smartphone was launched, thus reducing the number of irrelevant results. When searching ClinicalTrials.gov, the search was limited to studies first posted on or after January 1, 2008. Only studies published in English were included to ensure an accurate interpretation. Observational studies such as cross-sectional surveys, cohort studies, qualitative studies, economic studies, and intervention studies were included.

Intervention studies were not required to have a specific comparator or any comparators. Studies were excluded if they were solely descriptive of the app.

To understand the latest developments in accessible technology supporting improvement in the uptake of childhood vaccinations, we restricted this review to apps hosted on mobile platforms. The app could provide any service related to the promotion of vaccination or vaccination decision making, including but not limited to information sharing and record storing or sharing, and appointment support. Studies that did not involve the use or study of an app or solely focused on other ways of delivering vaccination interventions such as text messaging, telephone calls, or web-based interventions were excluded. Owing to the specific nature of the intervention, the population was restricted to children, parents, guardians, and/or health care professionals involved in the management of children. Children were defined as individuals aged less than or equal to 18 years. Studies focusing on the vaccination of adults were excluded. The study could have been conducted in any geographical setting.

### Outcome Measures

The primary outcome of this review was the uptake of vaccination. The secondary outcomes were the knowledge and decision making of parents; costs and cost-effectiveness; use of the app; measures of usability, for example, usefulness, acceptability, and experiences of different users (parents and health care professionals); and adverse events (eg, data leak and misinformation).

### Screening and Selection of Studies

All studies retrieved from the databases were stored in Mendeley version 1.19.5 (Elsevier), a reference management software. This software automatically eliminated duplicates before screening the citations against the inclusion and exclusion criteria by 2 independent reviewers. When duplicates, or publications from the same study were identified, the more recent publication or the one with the most details was selected for inclusion in the review. Any disagreements were discussed, and if a consensus was not reached, a third reviewer was consulted.

Published results of trials that were retrieved from CENTRAL or ClinicalTrials.gov and that met the inclusion criteria were searched for and included if not already captured; trial designs or protocols were excluded. A total of 10 trials met the inclusion criteria; 8 had no published data at the date of screening, and the published results of the remaining 2 trials were already included. The titles of references of 5 relevant review studies that were retrieved with our search strategy were reviewed for inclusion; 4 additional references were identified and were included in the full-text review.

The full text of the abstracts that met the inclusion criteria was screened by one of the reviewers and validated by a second reviewer to determine the studies to be included in the final set. Overall, 10 of the screened studies eligible for inclusion were conference or meeting abstracts and did not have full texts available; hence, they were excluded.

### Data Extraction

Data were extracted by 1 reviewer, and key data points from the studies that were specified in the protocol and identified on further study of the publications were recorded in a spreadsheet. The data extraction form was based on the minimum requirements recommended by the Cochrane Handbook for Systematic Reviews [24]. The data extracted from the studies are shown in Table 1. This process was validated by a second reviewer, and disagreements were resolved by a third reviewer. There were no disagreements in the extracted data; however, there were 18 instances in which the second reviewer proposed the inclusion of additional data for greater clarity.

**Table 1.** Data extracted from the included studies.

| Study information | Data extracted |
|---|---|
| General study information | Title of publication, year of publication, authors, and journal of publication |
| Study characteristics | Study design, country of study, analyzed sample size, key inclusion/exclusion criteria, and study arms |
| Intervention characteristics | App name, device on which the app could be or was utilized, compatible platforms, intended user, aim of the app, vaccines covered by the app, and vaccine-related features of the app |
| Evaluation | Number of users, impact on the uptake of vaccinations, impact on knowledge/learning, impact on vaccination decision making, perceived credibility, usability/user experiences, popular features, costs/cost-effectiveness, adverse events, and conclusions |

### Risk of Bias

The quality assessment of the included studies was undertaken by 1 reviewer and validated by a second reviewer. Any disagreements were resolved by consensus or the opinion of a third reviewer, where required. The methods specified in the Cochrane collaboration tool for assessing the risk of bias were used. The Cochrane collaboration risk of bias tool was used to assess the quality of the randomized controlled trials [25]; the risk of bias in nonrandomized studies of interventions (ROBINS-I) tool was used to assess the nonrandomized intervention trials [26]; and the critical appraisal skills program tools for cohort, qualitative, and economic studies were used for pre-post and quantitative studies, qualitative studies, and economic studies, respectively, [27-29]. The quality of cross-sectional survey studies was assessed using the Appraisal tool for Cross-Sectional Studies (AXIS) tool [30]. The results of the Cochrane collaboration risk of bias tool and ROBINS-I evaluations were summarized using RevMan 5.3 [31] and robvis [32], respectively. The critical appraisal skills program scores were calculated using standard practice, yes=1, no=0, and cannot tell=0 for each question, following which the total score was summed for each study. AXIS scores were summarized tabularly, and the mean and SD were calculated.

### Data Analysis and Synthesis

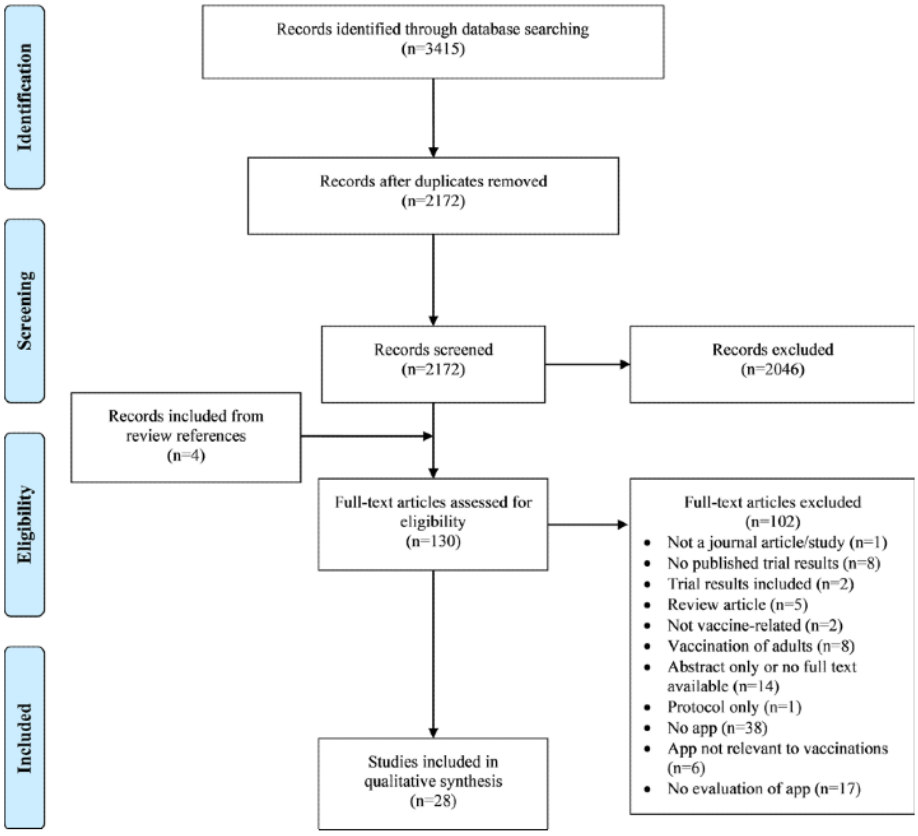
Owing to the variability in populations, interventions, outcomes, and study designs, a meta-analysis of the studies was not possible; hence, we report a narrative overview of the findings to draw conclusions about the potential roles, value, and effectiveness of the apps to support childhood vaccinations. For the purpose of this review, the app was considered to provide significant benefit if there was a statistically significant ($P \leq .054$) improvement in a given outcome as compared with a comparator or control or over time. If no significance was reported or if the difference was nonsignificant or significantly worse among groups or over time, the app was considered to have no significant evidence supporting it. The limitations and future directions for research were also summarized.

## Results

### Study Selection

Overall, 3415 studies were retrieved from the 7 databases; of these, 1243 were duplicates. Of the 2172 citations screened, 126 were selected for full-text review, and 4 additional studies were identified during the title screening of the references from 5 relevant review studies that had been retrieved in the database search. The primary reasons for exclusion at the screening stage were that the study was not vaccination-related (n=1171), did not include a mobile app (n=564), or was not health-related (n=89). Overall, 28 papers were included in the final review. The reasons for the exclusion of full-text review are detailed in Figure 1 [33].

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of study selection.




### Study Characteristics

The study characteristics of the 28 studies included in this review can be seen in Multimedia Appendix 2. These studies were published between 2010 and 2019. Overall, 3 of the 28 studies were randomized controlled trials [34-36], and 12 had a nonrandomized trial design; 9 were pre-post (before-and-after) studies [37-45], 2 were nonrandomized controlled trials [46,47], and 1 had an interrupted time series design [48]. A further 6 studies were cross-sectional survey studies [49-54], 4 were longitudinal observational studies [55-58], 2 were qualitative studies [59], and 1 was an economic design study [60].

The 28 studies evaluated 25 unique apps; 3 papers evaluated the ImmunizeCA app [39,58,59] and 2 studies assessed MorbiQuiz [35,49]. The included studies varied in design, population, and geographical setting. The study populations generally included parents (15/28) [35-39,44,46-52,54,59], expectant mothers (5/28) [36,39,44,47,59], and/or children (4/28) [41-43,61]. Other populations included the general public (5/28) [45,53,55,57,58], households (1/28) [43], and village doctors (1/28) [34,43]. Some studies included multiple populations; hence, they were included multiple times in these statistics. The population size in the included studies ranged from 6 to 161,695 participants [43,51]. Overall, 12 of the 28 included studies were conducted in North American countries [37-40,44,45,51,53,57-60], 7 in Asia [34,36,41,42,47,52,56], 1 in the Middle East [54], 5 in Europe [35,46,49,50,61], 2 in Africa [43,48], and 1 included worldwide users [55]. Approximately, 39% (11/28) of the studies took place in

deprived areas and/or developing countries [34,36,38,41-43,47,48,52,54,56].

## App Characteristics

The characteristics of the 25 apps investigated in the included studies are shown in Table 2. The apps were primarily delivered via a smartphone or tablet, with 4 apps using multimodal delivery methods [36,37,50,56]. Over 50% (15/25) of the apps were intended for use by parents [35,37,38,40,44,46,50-52,54-58,60], with 4 designed specifically for use by mothers [44,52,54,58]. Furthermore, 3 apps were for use by multidisciplinary populations involved in the delivery of childhood vaccinations, for example, health care providers, pharmacists, and parents [50,55,56]. Overall, 7 of the 25 apps were designed solely for the use of health care providers/health workers [34,36,41-43,47,48]. The vaccines covered by these apps varied. Almost two-thirds (15/25) of the apps covered multiple vaccines [34,36,41,42,44,46-48,50-52,54-58], 5 focused solely on HPV [37,38,40,60,61], 2 on influenza [45,53], 1 on measles-mumps-rubella [35], and 1 on measles only [43]. Furthermore, 4 studies describing 4 unique apps reported an association with the Expanded Program on Immunization (EPI) [34,41,42,47]. All 25 apps served vaccination-related functions; however, 28% (7/25) had a focus beyond vaccinations [36,42,45,50,52,54,56]. The scope of these apps remained broadly within the field of antenatal, maternal, or child health, with the exception of the Carrot Rewards app, which was intended for general health initiatives [45].

**Table 2.** Characteristics of childhood vaccination apps.

| App | Intended user | Technical specifications | Compatible platforms |
|---|---|---|---|
| Conversational agent for HPV[a] vaccination [37] | Parent | • Graphical user interface tool on laptop used by an operator<br>• Bluetooth communication with iPad, the user-facing interface<br>• Text-to-speech capability<br>• Wizard of Oz agent architecture | Apple iOS[b] and MacOS |
| ImmunizeCA [39,58,59] | Women of childbearing age | • Generates customized vaccination schedules<br>• Vaccine information available<br>• Creates virtual immunization record<br>• Syncs with calendar for scheduling<br>• Embedded outbreak alert feature<br>• Basic security features<br>• Rotating banner in app used to display features and public health messages | iOS and Android |
| Tablet-based self-persuasion app [38] | Parent | • Voiceover narration of task<br>• Audio recording function to facilitate self-administration | iOS |
| ReadyVax [38,55] | Health care providers, pharmacists, parents, and patients | • Native app direct to smartphone<br>• Offline functioning<br>• Information updates automatically<br>• Browsable and searchable information<br>• Information updated through a web-based dashboard interface<br>• Alert notifications can be sent<br>• Links to multimedia | iOS |
| UberHealth [53] | Anyone | • Request and delivery of vaccines using geolocation software | NR[c] |
| EPI[d] app [34] | Doctors | • Record vaccination status and upload data into the CIRS[e]<br>• CIRS sends daily updates on children for whom vaccination is overdue<br>• Contact details of families available | NR |
| Carrot Rewards [45] | Anyone | • In-app quiz about influenza vaccinations<br>• Geolocation-based push notification when in proximity to a pharmacist<br>• Loyalty points for completion of vaccination-related tasks | NR |
| MorbiQuiz [35,49] | Parent | • Daily quiz targeting vaccination literacy<br>• Vaccination empowerment videos<br>• Leaderboard for quiz results | iOS and Android |
| Tablet-based HPV educational module [40] | Patient and parent | • Educational videos on HPV<br>• Flashcard information on HPV | NR |
| RapidSMS [48] | Health worker | • Mobile alert system for vaccination tracking | NR |
| CHeITA [50] | Health care providers, parents, and guardians | • Stores health history of family and development statistics<br>• Vaccination tracking | Windows, iOS, and Android |

| App | Intended user | Technical specifications | Compatible platforms |
|---|---|---|---|
| Mobile technology supporting EPI coverage [41] | Health worker | • Stores personal and familial information<br>• Case identification via pictures<br>• Pronunciation of the name of the child in the mother's ethnic language | NR |
| Mother and Child Care Module-EPI module [42] | Health worker | • Immunization status collected<br>• Connection with server module<br>• Generates appointment dates and SMS reminders | NR |
| Tailored interactive multimedia intervention [60] | Parent | • Tailored interactive health communication about HPV via videos | NR |
| Baby Care app [52] | Mothers | • Embedded FAQs[f]<br>• Upload child data<br>• Trend analysis<br>• Alert messaging<br>• Baby's periodic health report generation | NR |
| EpiSurveyor [43] | Health worker | • Sources of information<br>• Basic demographics<br>• Consent to bring children for immunization | Android |
| ImTeCHO [36] | Health worker | • Registration of pregnant women and children aged under 2 years<br>• Generates daily appointment schedule<br>• Videos to emphasize key health messages | Android and web |
| Smartphone App for Premature Infants [54] | Mothers | • Electronic learning modules | NR |
| Call the shots [51] | Parent | • Reminders for vaccination<br>• Record keeping of child's vaccinations<br>• Hosts latest immunization schedule<br>• Extensive toolkit embedded with FAQs<br>• Links to videos and resources | Android |
| FightHPV [61] | Teenagers | • Gamified narratives with connected text messages to convey HPV information<br>• Players able to share information with social network | iOS, Java, and Android |
| MomsTalkShots [44] | Mothers | • Videos with obstetricians and pediatricians of different ethnicities<br>• Intervention tailored to knowledge and beliefs | NR |
| VaccApp [46] | Parent | • Avatar requests vaccination information | Android |
| iCHRcloud [56] | Parent and doctor | • Mobile interface, doctor module, and cloud component<br>• Child health records stored, updated, and shared across network | iOS and Android |
| mTika [47] | Health worker | | Android |

| App | Intended user | Technical specifications | Compatible platforms |
|---|---|---|---|
| | | • Registration of pregnant women<br>• SMS birth notifications from mothers<br>• Automated SMS vaccination reminders to mothers and health workers<br>• EPI monitoring by supervisors | |
| iPhone app [57] | Parent | • Stores child's vaccination information<br>• Hosts recommended vaccination schedule<br>• Generates customized vaccination schedule | NR |

[a]HPV: human papillomavirus.

[b]iOS: iPhone operating system.

[c]NR: not reported.

[d]EPI: Expanded Program on Immunization.

[e]CIRS: Child Immunization Register System.

[f]FAQs: frequently asked questions.

## Functionality of Childhood Vaccination Apps

The investigated apps were most commonly designed for the primary purpose of education (11/25) [35,40,43-46,52,54,55,60,61], record keeping (8/25) [34,36,42,47,50,56-58], or reminder systems (3/25) [41,48,51], as shown in Table 3. Despite specifying distinct primary functions, the apps had multiple overlapping capabilities. On average, the apps performed 1.9 functions (range 1-4), with the most commonly occurring functions aligning with the primary functions: education (14/25), management of records (12/25), and reminders (11/25).

There was no consistent reporting of the most popular features, perceptions, or usage of individual functions. A total of 5 studies reporting on the usage of the apps noted that the most commonly used/most popular features were those that helped manage vaccination records [39], provided vaccination information [55], supported appointment management [34,59], and generated summary reports [52]. These functions aligned with the primary functions of the app being investigated. A qualitative study supplemented with Google Analytics data for the ImmunizeCA app reflected the overall data. These researchers reported that 9 of the 10 women interviewed used the vaccination tracking function, 80% used the appointment reminders/calendar, and 80% used the information on vaccines [59]. This was echoed by the Google Analytics data that were reported in the same study, wherein 47.6% of all app sessions accessed the tracking function, compared with 9.5% and 4.9%, where the appointment reminders and vaccination information were accessed, respectively [59]. One study that asked participants about the helpfulness of specific features of the app reported that all 6 respondents found the date reminder system helpful, whereas 5 respondents found the vaccine information very helpful [51].

**Table 3.** Capabilities of the apps described in the included studies.

| App | Counseling | Self-persuasion | Management of records | Reminders | Vaccine-preventable disease breakout alert | Education | Frequently asked questions | Vaccine delivery | Total |
|---|---|---|---|---|---|---|---|---|---|
| Conversational agent for HPV[a] vaccination [37] | X[b] | N/A[c] | N/A | N/A | N/A | N/A | N/A | N/A | 1 |
| Immunize-CA [39,58,59] | N/A | N/A | X[b] | X[d] | X[d] | N/A | X[d] | N/A | 4 |
| Tablet-based self-persuasion app [38] | N/A | X[b] | N/A | N/A | N/A | X[d] | N/A | N/A | 2 |
| ReadyVax [55] | N/A | N/A | N/A | N/A | X[d] | X[b] | X[d] | N/A | 3 |
| UberHealth [53] | N/A | N/A | N/A | N/A | N/A | N/A | N/A | X[b] | 1 |
| EPI[e] app [34] | N/A | N/A | X[b] | X[d] | N/A | X[d] | N/A | N/A | 3 |
| Carrot Rewards [45] | N/A | N/A | N/A | X[d] | N/A | X[b] | N/A | N/A | 2 |
| MorbiQuiz [35,49] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |
| Tablet-based HPV educational module [40] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |
| RapidSMS [48] | N/A | N/A | N/A | X[b] | N/A | N/A | N/A | N/A | 1 |
| CHeITA [50] | N/A | N/A | X[b] | N/A | N/A | N/A | N/A | N/A | 1 |
| Mobile technology supporting EPI coverage [41] | N/A | N/A | X[d] | X[b] | N/A | X[d] | N/A | N/A | 3 |
| Mother and child care module-EPI module [42] | N/A | N/A | X[b] | X[d] | N/A | N/A | N/A | N/A | 2 |
| Tailored interactive multimedia intervention [60] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |
| Baby Care app [52] | N/A | N/A | X[d] | X[d] | N/A | X[b] | X[d] | N/A | 4 |
| EpiSurveyor [43] | N/A | N/A | X[d] | N/A | N/A | X[b] | N/A | N/A | 2 |
| ImTeCHO [36] | N/A | N/A | X[b] | N/A | N/A | N/A | N/A | N/A | 1 |
| Smartphone App for Premature Infants [54] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |

| App | Counseling | Self-persuasion | Management of records | Reminders | Vaccine-preventable disease breakout alert | Education | Frequently asked questions | Vaccine delivery | Total |
|---|---|---|---|---|---|---|---|---|---|
| Call the shots [51] | N/A | N/A | X[d] | X[b] | N/A | N/A | X[d] | N/A | 3 |
| FightHPV [61] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |
| MomsTalk-Shots [44] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |
| VaccApp [46] | N/A | N/A | N/A | N/A | N/A | X[b] | N/A | N/A | 1 |
| iCHRcloud [56] | N/A | N/A | X[b] | X[d] | N/A | N/A | N/A | N/A | 2 |
| mTika [47] | N/A | N/A | X[b] | X[d] | N/A | N/A | N/A | N/A | 2 |
| iPhone app [57] | N/A | N/A | X[b] | X[d] | X[d] | N/A | N/A | N/A | 3 |
| Total | 1 | 1 | 12 | 11 | 3 | 14 | 3 | 1 | N/A |

[a]HPV: human papillomavirus.

[b]X: indicates primary functions.

[c]N/A: not applicable.

[d]X: indicates secondary functions.

[e]EPI: Expanded Program on Immunization.

### Uptake of Vaccinations

The extracted outcomes and results are provided in Multimedia Appendix 3. Overall, 9 of the 28 studies assessed the impact of an app on vaccination uptake. Furthermore, 4 studies reported a significant improvement in vaccination coverage after versus before the implementation of the app [34,42,45,47]. These studies reported a 17% ($P$=.03) [34], 5% ($P$<.001) [45], 9.7% ($P$<.001) [42], and 17.9% (rural) and 16.4% (urban; $P$<.001 for both) [42,47] increase in the vaccination rates after versus before the implementation of the app. In addition, 2 of the 4 studies included a control group. The study of the mTika app identified a significant difference-in-difference estimate of the difference between the intervention groups' change from baseline to end line and the control groups' change from baseline to end line of fully vaccinated children (21.6% rural and 23.2% urban difference; $P$<.05) [47]. Conversely, the study of the EPI app did not find a significant difference between the intervention and control groups at the end line (2.5% difference; $P$=.16).

Of the remaining 5 studies, 2 reported no significant benefit [36,48] and 3 reported no significance level [41,43,53] regarding an increase in the vaccination rates between the intervention and control groups or after versus before intervention implementation.

### Vaccination Knowledge and Decision Making

A total of 10 studies reported on the impact of the vaccination apps on knowledge/learning, as shown in Multimedia Appendix 3. Furthermore, 4 studies reported significant improvements in the knowledge or learning compared with a control group or after versus before the intervention ($P$≤.05) [35,40,46,61], 2 reported no significant improvement [39,41], and 4 did not indicate a significance level [43,47,49,54].

The implications of the vaccination apps on decision making and evaluation of the risk-benefit of vaccinations were investigated in 8 studies, as shown in Multimedia Appendix 3. Furthermore, 7 of the 8 studies reporting this outcome indicated a positive impact of the apps on vaccination beliefs and intent to vaccinate [35,38-41,44,49]. The remaining study was unable to report improvements because of no vaccine hesitancy in participants at baseline [37]. Half of the studies (4/8) indicated significant improvements ($P$≤.05) in the intent to vaccinate children, positive attitudes toward vaccination, and/or confidence in their vaccination decision after interaction with the app versus before or compared with a control group [35,38,40,41]. The remaining 4 studies did not report a significance level [37,39,44,49]. Fadda et al [35] investigated the knowledge and empowerment functions separately and found that exposure to the knowledge intervention significantly improved the intent to vaccinate ($P$=.03) and confidence in the participants' vaccination decision ($P$=.006) versus control, but empowerment and combined interventions did not show stronger intent to vaccinate. Furthermore, 3 studies indicated that as well as having the potential to promote vaccination, the apps also had the potential to discourage users from vaccinating their children [39,44,49].

### Costs/Cost-Effectiveness

Only 1 study reported on the costs or cost-effectiveness of a childhood vaccination app. The cost of developing a computerized, tailored, interactive multimedia intervention was found to be approximately double the cost of a print-based Photonovella intervention for HPV vaccine education (US

$135,978 vs US $66,468, respectively). This difference was retained in amortized annual costs over a 7-year period (US $21,825 vs US $10,669 per year for the tailored, interactive multimedia intervention and Photonovella, respectively) [60].

### Usability and Acceptability

Overall, 9 studies reported on the usability/ease of use (n=5), acceptability (n=1), or both (n=3) aspects of the vaccination apps. Furthermore, 8 of these studies reported high ease of use (average score for ease of use/usability >70%, or >70% of the participants rated the app easy to use). A total of 2 studies also reported high acceptance of the app (average score for acceptance >70%, or >70% of participants reporting acceptance).

### Participant Perceptions

A total of 11 studies reported on participants' perceptions of childhood vaccination apps. Furthermore, 9 studies reported on the perceptions of parents [38-40,44,46,49,51,52,54], 1 reported on teenagers' experiences [61], and 1 reported on the perceptions of mothers and vaccination service providers [47]. All 10 studies reporting quantitative results indicated positive user experiences, with participants considering the app to be helpful and/or trustworthy or reporting that they were satisfied, confident, and/or likely to recommend the app (average score of >70%, or >70% of the participants agreeing with relevant statements) [38-40,44,46,49,51,52,54,61]. The study reporting on the qualitative experiences of service providers and mothers with the mTika app revealed that the app was perceived as helpful, easily understood by mothers, user-friendly, time-efficient, and helpful in reducing the workload of vaccination service providers [47].

### Risk of Bias Assessment

Owing to the heterogeneity of the study types, a variety of quality assessment tools were employed to assess the risk of bias for the 28 included studies. The summary tables and figures are provided in Multimedia Appendix 4. Overall, the quality of the studies assessed in this review ranged from moderate to poor. The studies assessed using the critical appraisal skills program cohort, qualitative, and economic checklists met on average 6.4 out of 12 (range 4-9), 6.5 out of 10 (range 6-7), and 9 out of 12 (range 9) criteria, respectively. Cross-sectional studies assessed using the AXIS tool had a mean score of 9.3 (SD 2.2) out of 20. Over 50% (15/28) of the studies were deemed to have inappropriate recruitment strategies [34,36,37,41,42,45,46,48,49,52,54,56,58,61], primarily owing to the lack of sufficient information. Risk of bias in exposure (performance bias) was identified in 13 out of 17 studies in which this was assessed [34-37,41,43-45,48,55-58]. Outcome bias (detection bias) was suspected in 5 out of 19 studies in which it was assessed [34,36,41,42,48]. The implications/value of the research and fit of the results in context was lacking in 11 out of 17 [37,39,42,44,48,55-60] and 7 out of 15 [37,40,43-45,56,60] studies assessed for these criteria, respectively.

The critical appraisal skills program cohort checklist assessed confounding, completeness, and the duration of follow-up. The identification and mitigation of confounders was not found to be sufficient in any of the 14 studies assessed using this checklist [37-42,44,45,48,55-58]. A total of 13 studies gave no information on confounders, and the remaining study noted that despite the identification of potential confounders, they were unable to perform fixed or random-effects modeling to assess the impact of these factors [38]. Data on dropout rate were not recorded systematically (2/14) [45,58]; this was primarily because of a lack of substantiation of follow-up duration, information on the timing of follow-up, and/or no indication of whether discontinuations were significant.

Six cross-sectional studies assessed using the AXIS tool were found to lack justification of sample size, definition of target population (1/6), categorization of (1/6) [50,51] and information on (0/6) nonresponders, validation of outcome measures (2/6) [52,54], repeatability (2/6) [49,51], and internal consistency of results (2/6) [51]. Most of these shortcomings were because of a lack of information. Moreover, 1 of the 2 nonrandomized controlled studies was deemed to have an overall serious risk of bias because of the serious risk of selection bias in determining the intervention groups [46].

## Discussion

### Principal Findings

In this systematic review, 28 studies evaluating 25 childhood vaccination apps were examined. Overall, there is little evidence to suggest that childhood vaccination apps are effective in improving vaccination coverage, with only 4 of the 9 studies assessing this outcome indicating significant benefit ($P \leq .05$) after versus before the app was introduced. This contrasts with a systematic review and meta-analysis conducted by Harvey et al [14], which revealed a significant benefit of reminder ($P < .001$), recall ($P = .02$), reminder and recall ($P < .006$), and educational initiatives on childhood vaccination rate ($P = .02$); however, these prompts were not delivered via the mobile app. Our findings also contrast with systematic reviews that have found mobile apps to be effective in eliciting health-related behavior change [62,63].

Similarly, 4 out of 10 studies assessing the impact on vaccination knowledge and 4 out of 8 studies assessing vaccination decision making reported significant benefit of the app ($P \leq .05$). Furthermore, 3 studies substantiated the dichotomous effect of information provision, thus illustrating the potential for apps to dissuade individuals from vaccination. This is in keeping with evidence from other studies [64,65]. Parental decision making regarding childhood vaccination is understood to be a specific scenario for health-related decision making where parents have been stipulated to put major weight on the subjective perception of the outcome [66]. It is, therefore, important to understand the likely interpretation of the information provided and how this may affect the parental risk-benefit analysis of vaccination.

The primary functionality of the apps described in the included studies varied; however, most had multiple functions, with the most common features being education, reminders, and record keeping. These were primarily for the use of parents or health care providers; only the iCHRcloud app facilitated the sharing of vaccination record information between parents and physicians [56]. No apps had been designed for the use of school

staff, and no apps had the functionality to share information between schools and parents. In the United Kingdom, many childhood and adolescent vaccinations are delivered at schools for convenience and to enhance delivery [67,68]. This could be an avenue for further investigation.

There is insufficient evidence to draw any conclusions regarding the relationship between the function of the app and the efficacy in improving vaccination rates, knowledge, or positive decision making. The 25 investigated apps had diverse functionalities but were primarily designed for providing vaccine information and/or record keeping. Studies reporting on user statistics revealed that the most popular functions were record keeping, reminders, and information access. Overall, usability, acceptability, and user perceptions of the apps were positive.

### Quality of the Evidence

The quality assessment of the included studies revealed that many were of poor to moderate quality, indicating an overall high risk of bias, which risks impairing the validity of the conclusions regarding the effectiveness of childhood vaccination apps. One study was determined to have a serious risk of bias. Most negative indicators were because of a lack of information about the criteria assessed. The risk of bias and inadequate robustness may be because of the nature of many of the included studies being pilot, early usability, and preliminary scoping studies. To draw valid and accurate conclusions on the quality of studies, study methods should be comprehensively reported. The infancy of these types of apps also had an impact on the assessment of the implications/value and the fit of the results in context, as many studies indicated that they were the first of their kind in their setting.

### Strengths and Weaknesses of the Study

The strengths of this study lie in the comprehensive analysis of the available literature discussing apps for childhood vaccination. We investigated ClinicalTrials.gov and ERIC databases, which include gray literature, and we included letters and full-text conference proceedings [50,53]. The inclusion of gray literature minimizes publication bias and ensures that a comprehensive view of the latest literature is reported [69]. We also included all study types, thereby ensuring that apps at every stage of development were reported; a study limited to randomized controlled trials may omit studies of apps in their infancy. One limitation of this review is because of the heterogeneity of the studies and their reported outcomes; therefore, it was not possible to conduct a meta-analysis. A meta-analysis would enable the quantification of the heterogeneity of studies and allow us to quantify the effectiveness of childhood vaccination apps.

### Implications of the Study

Immunization is a simple and effective mechanism for reducing childhood mortality. Despite the insignificant findings of this review about the effect of apps on the uptake of vaccinations, the positive user perceptions, usability, and acceptability reported present a compelling opportunity to build on the successes of current apps and learn from their shortcomings. Individual studies included in this review reported the potential benefit of these apps on an individual, community, and nationwide level, highlighting the breadth of engagement that can be harnessed with the use of mobile apps [42,45,47]. It is evident that more needs to be done to ensure that a positive change in vaccination knowledge and decision making is paralleled with the provision of sufficient, accessible resources to allow vaccinations to be easily obtained. There are important aspects to consider with regard to the needs of users and how apps will be implemented in health care service delivery.

### Unanswered Questions and Future Research

Mobile apps will likely play a role in the storage and sharing of vaccination records, generation of reminders, and/or dissemination of vaccination education. Despite several publicly available apps and others in development, a lack of robust evidence remains regarding the effectiveness of vaccination apps in improving vaccination coverage. This systematic review, despite not reporting significant efficacy, indicates that many of the apps convey some degree of benefit with regard to improving vaccination uptake, knowledge, and/or decision making and are widely accepted by their users. For future research, it will be important to understand the priorities of different user groups in terms of app functionalities and the dichotomous effects of vaccination information. Many of the studies included in this review are early-stage investigations and were found to have a relatively high risk of bias. Further investigation of these apps in larger, more robust, controlled trials will allow greater granularity of evaluation and understanding of the role and implications of these apps for wider communities and various subpopulations. In addition, many of the included studies originated from developing countries; however, preventable childhood illnesses are increasing globally. Outcomes from similar studies in developed countries may present a different picture.

## Conclusions

The objective of this systematic review was to evaluate the effectiveness of childhood vaccinations in improving vaccination uptake, knowledge, and decision making as well as to investigate the usability and patient perceptions of these interventions. Overall, 28 studies describing 25 apps were investigated. Although the apps were generally positively received and had high usability and acceptability scores, there was little evidence to suggest that they were effective in significantly improving uptake, knowledge, or decision making; however, most apps were seen to provide some benefit. This indicates that there is demand and engagement with apps supporting childhood vaccinations; however, further investigation is required.

The studies investigating these apps were considered to be of poor to moderate quality, likely because of the early phase nature of many of the apps and their respective studies. Only 5 studies were randomized. An additional concern raised by 3 studies was the potential for these apps to discourage vaccination among those who were initially undecided about infant vaccination and among those who had previously intended to vaccinate their children. Future research is warranted into the dichotomous effects of the provision of vaccination information, the outcomes of larger robust studies of these apps, and the needs and priorities of various user populations.

## Acknowledgments

CC, EM, MI, and MV were supported in this study by EIT Health (grant 18654).

## Authors' Contributions

MM, MV, and EM conceived the study topic and designed the review protocol. CC and MI screened the studies. CC conducted the data extraction and risk of bias assessment, which were validated by MI. The systematic review was written by CC with revisions from MV, EM, MM, and MI.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 14 KB - mhealth_v8i5e17371_app1.docx ]

Multimedia Appendix 2
Study characteristics.
[DOCX File , 18 KB - mhealth_v8i5e17371_app2.docx ]

Multimedia Appendix 3
Study outcomes and extracted effectiveness results.
[DOCX File , 20 KB - mhealth_v8i5e17371_app3.docx ]

Multimedia Appendix 4
Risk of bias assessment.
[DOCX File , 1299 KB - mhealth_v8i5e17371_app4.docx ]

## References

1. World Health Organization. 2019. Immunization Coverage URL: https://www.who.int/news-room/fact-sheets/detail/immunization-coverage [accessed 2019-11-25]
2. Hotez P. America and Europe's new normal: the return of vaccine-preventable diseases. Pediatr Res 2019 Jun;85(7):912-914. [doi: 10.1038/s41390-019-0354-3] [Medline: 30812027]
3. Tanne JH. Measles: two US outbreaks are blamed on low vaccination rates. Br Med J 2019 Jan 21;364:l312. [doi: 10.1136/bmj.l312] [Medline: 30665901]
4. NHS Digital. Childhood Vaccination Coverage Statistics-England 2018-19 URL: https://digital.nhs.uk/data-and-information/publications/statistical/nhs-immunisation-statistics/england-2018-19 [accessed 2019-11-25]
5. Hill HA, Elam-Evans LD, Yankey D, Singleton JA, Kang Y. Vaccination coverage among children aged 19-35 months-United States, 2017. MMWR Morb Mortal Wkly Rep 2018 Oct 12;67(40):1123-1128 [FREE Full text] [doi: 10.15585/mmwr.mm6740a4] [Medline: 30307907]
6. McKee C, Bohannon K. Exploring the reasons behind parental refusal of vaccines. J Pediatr Pharmacol Ther 2016;21(2):104-109 [FREE Full text] [doi: 10.5863/1551-6776-21.2.104] [Medline: 27199617]
7. ScienceDaily. Vaccinations of US Children Declined After Publication of Now-Refuted Autism Risk URL: https://www.sciencedaily.com/releases/2012/06/120604142726.htm [accessed 2019-11-25]
8. Graham JE, Mishra A. Global challenges of implementing human papillomavirus vaccines. Int J Equity Health 2011 Jul 30;10:27 [FREE Full text] [doi: 10.1186/1475-9276-10-27] [Medline: 21718495]
9. Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. Vaccine 2014 May 17;32(19):2150-2159. [doi: 10.1016/j.vaccine.2014.01.081] [Medline: 24598724]
10. Greenlee CJ, Newton SS. A review of traditional vaccine-preventable diseases and the potential impact on the otolaryngologist. Int Arch Otorhinolaryngol 2018 Jul;22(3):317-329 [FREE Full text] [doi: 10.1055/s-0037-1604055] [Medline: 29983775]
11. Merrill R. Introduction to Epidemiology. Burlington, Massachusetts: Jones & Bartlett Learning; 2012.
12. Nelson KN, Wallace AS, Sodha SV, Daniels D, Dietz V. Assessing strategies for increasing urban routine immunization coverage of childhood vaccines in low and middle-income countries: a systematic review of peer-reviewed literature. Vaccine 2016 Nov 4;34(46):5495-5503 [FREE Full text] [doi: 10.1016/j.vaccine.2016.09.038] [Medline: 27692772]

13. Walling EB, Benzoni N, Dornfeld J, Bhandari R, Sisk BA, Garbutt J, et al. Interventions to improve HPV vaccine uptake: a systematic review. Pediatrics 2016 Jul;138(1):e20153863 [FREE Full text] [doi: 10.1542/peds.2015-3863] [Medline: 27296865]
14. Harvey H, Reissland N, Mason J. Parental reminder, recall and educational interventions to improve early childhood immunisation uptake: a systematic review and meta-analysis. Vaccine 2015 Jul 9;33(25):2862-2880 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.085] [Medline: 25944299]
15. Springfield Centre. Springfield Centre. 2012. Making ICT Work For Bangladesh's Farmers URL: https://www.springfieldcentre.com/making-ict-work-for-bangladeshs-farmers/ [accessed 2019-11-25]
16. World Health Organization. mHealth New Horizons for Health Through Mobile Technologies: Based on the Findings of the Second Global Survey on eHealth URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2019-11-25]
17. Google Play. Vaccination Apps URL: https://play.google.com/store/search?q=vaccination&c=apps&gl=GB [accessed 2019-11-25]
18. Abahussin AA, Albarrak AI. Vaccination adherence: review and proposed model. J Infect Public Health 2016;9(6):781-789 [FREE Full text] [doi: 10.1016/j.jiph.2016.09.006] [Medline: 27659114]
19. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey: Wiley; 2008.
20. Centre for Reviews and Dissemination. Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care. York, England: University of York; 2009.
21. van Velthoven M, Milne-Ives M, de Cock C, Mooney M, Meinert E. Use of apps to promote childhood vaccination: a systematic review protocol. JMIR Res Protoc 2019:- epub ahead of print. [doi: 10.2196/preprints.16929]
22. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jul 15;7:16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
23. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J 2015 Jan 2;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
24. Higgins JP. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: Wiley; 2008.
25. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
26. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. Br Med J 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
27. Critical Appraisal Skills Programme. CASP Checklist: 12 Questions to Help You Make Sense of a Cohort Study URL: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Cohort-Study-Checklist-2018_fillable_form.pdf [accessed 2019-11-20]
28. Critical Appraisal Skills Programme. CASP Checklist: 10 Questions to Help You Make Sense of a Qualitative Research URL: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Qualitative-Checklist-2018_fillable_form.pdf [accessed 2019-11-20]
29. Critical Appraisal Skills Programme. CASP Checklist: 12 Questions to Help You Make Sense of an Economic Evaluation URL: https://casp-uk.net/wp-content/uploads/2018/03/CASP-Economic-Evaluation-Checklist-2018_fillable_form.pdf [accessed 2019-11-20]
30. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016 Dec 8;6(12):e011458 [FREE Full text] [doi: 10.1136/bmjopen-2016-011458] [Medline: 27932337]
31. Cochrane Training. 2014. Cochrane Review Manager (RevMan) URL: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman [accessed 2020-04-28]
32. GitHub. A Package to Quickly Visualise Risk-of-Bias Assessment Results URL: https://github.com/mcguinlu/robvis [accessed 2019-11-24]
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
34. Chen L, Du X, Zhang L, van Velthoven MH, Wu Q, Yang R, et al. Effectiveness of a smartphone app on improving immunization of children in rural Sichuan province, China: a cluster randomized controlled trial. BMC Public Health 2016 Aug 31;16:909 [FREE Full text] [doi: 10.1186/s12889-016-3549-0] [Medline: 27581655]
35. Fadda M, Galimberti E, Fiordelli M, Romanò L, Zanetti A, Schulz PJ. Effectiveness of a smartphone app to increase parents' knowledge and empowerment in the MMR vaccination decision: a randomized controlled trial. Hum Vaccin Immunother 2017 Nov 2;13(11):2512-2521 [FREE Full text] [doi: 10.1080/21645515.2017.1360456] [Medline: 29125783]

36. Modi D, Dholakia N, Gopalan R, Venkatraman S, Dave K, Shah S, et al. mHealth intervention 'ImTeCHO' to improve delivery of maternal, neonatal, and child care services-a cluster-randomized trial in tribal areas of Gujarat, India. PLoS Med 2019 Oct;16(10):e1002939 [FREE Full text] [doi: 10.1371/journal.pmed.1002939] [Medline: 31647821]
37. Amith M, Zhu A, Cunningham R, Lin R, Savas L, Shay L, et al. Early usability assessment of a conversational agent for HPV vaccination. Stud Health Technol Inform 2019;257:17-23 [FREE Full text] [Medline: 30741166]
38. Baldwin AS, Denman DC, Sala M, Marks EG, Shay LA, Fuller S, et al. Translating self-persuasion into an adolescent HPV vaccine promotion intervention for parents attending safety-net clinics. Patient Educ Couns 2017 Apr;100(4):736-741 [FREE Full text] [doi: 10.1016/j.pec.2016.11.014] [Medline: 27912928]
39. Atkinson KM, Westeinde J, Ducharme R, Wilson SE, Deeks SL, Crowcroft N, et al. Can mobile technologies improve on-time vaccination? A study piloting maternal use of ImmunizeCA, a pan-Canadian immunization app. Hum Vaccin Immunother 2016 Oct 2;12(10):2654-2661 [FREE Full text] [doi: 10.1080/21645515.2016.1194146] [Medline: 27322109]
40. Gockley AA, Pena N, Vitonis A, Welch K, Duffey-Lind EC, Feldman S. Tablet-based patient education regarding human papillomavirus vaccination in colposcopy clinic. J Low Genit Tract Dis 2019 Jul;23(3):188-192. [doi: 10.1097/LGT.0000000000000474] [Medline: 30973442]
41. Kaewkungwal J, Apidechkul T, Jandee K, Khamsiriwatchara A, Lawpoolsri S, Sawang S, et al. Application of mobile technology for improving expanded program on immunization among highland minority and stateless populations in northern Thailand border. JMIR Mhealth Uhealth 2015 Jan 14;3(1):e4 [FREE Full text] [doi: 10.2196/mhealth.3704] [Medline: 25589367]
42. Kaewkungwal J, Singhasivanon P, Khamsiriwatchara A, Sawang S, Meankaew P, Wechsart A. Application of smart phone in 'Better Border Healthcare Program': a module for mother and child care. BMC Med Inform Decis Mak 2010 Dec 3;10:69 [FREE Full text] [doi: 10.1186/1472-6947-10-69] [Medline: 21047412]
43. Mbabazi WB, Tabu CW, Chemirmir C, Kisia J, Ali N, Corkum MG, et al. Innovations in communication technologies for measles supplemental immunization activities: lessons from Kenya measles vaccination campaign, November 2012. Health Policy Plan 2015 Jul;30(5):638-644 [FREE Full text] [doi: 10.1093/heapol/czu042] [Medline: 24920218]
44. Salmon DA, Limaye RJ, Dudley MZ, Oloko OK, Church-Balin C, Ellingson MK, et al. MomsTalkShots: an individually tailored educational application for maternal and infant vaccines. Vaccine 2019 Oct 8;37(43):6478-6485. [doi: 10.1016/j.vaccine.2019.08.080] [Medline: 31506192]
45. Dale LP, White L, Mitchell M, Faulkner G. Smartphone app uses loyalty point incentives and push notifications to encourage influenza vaccine uptake. Vaccine 2019 Jul 26;37(32):4594-4600. [doi: 10.1016/j.vaccine.2018.04.018] [Medline: 29699784]
46. Seeber L, Conrad T, Hoppe C, Obermeier P, Chen X, Karsch K, et al. Educating parents about the vaccination status of their children: a user-centered mobile application. Prev Med Rep 2017 Mar;5:241-250 [FREE Full text] [doi: 10.1016/j.pmedr.2017.01.002] [Medline: 28127527]
47. Uddin MJ, Shamsuzzaman M, Horng L, Labrique A, Vasudevan L, Zeller K, et al. Use of mobile phones for improving vaccination coverage among children living in rural hard-to-reach areas and urban streets of Bangladesh. Vaccine 2016 Jan 4;34(2):276-283 [FREE Full text] [doi: 10.1016/j.vaccine.2015.11.024] [Medline: 26647290]
48. Hategeka C, Ruton H, Law MR. Effect of a community health worker mhealth monitoring system on uptake of maternal and newborn health services in Rwanda. Glob Health Res Policy 2019;4:8 [FREE Full text] [doi: 10.1186/s41256-019-0098-y] [Medline: 30949586]
49. Fadda M, Galimberti E, Fiordelli M, Schulz PJ. Evaluation of a mobile phone-based intervention to increase parents' knowledge about the measles-mumps-rubella vaccination and their psychological empowerment: mixed-method approach. JMIR Mhealth Uhealth 2018 Mar 7;6(3):e59 [FREE Full text] [doi: 10.2196/mhealth.8263] [Medline: 29514772]
50. Devos J, de Haes S. Ghent University Academic Bibliography. 2014. The Proceedings of The 8th European Conference on IS Management and Evaluation ECIME 2014 URL: https://biblio.ugent.be/publication/5697112/file/5697113.pdf [accessed 2019-11-21]
51. Peck JL, Stanton M, Reynolds GE. Smartphone preventive health care: parental use of an immunization reminder system. J Pediatr Health Care 2014;28(1):35-42. [doi: 10.1016/j.pedhc.2012.09.005] [Medline: 23195652]
52. Kuo M, Lu Y, Chang P. A newborn baby care support app and system for mhealth. NI 2012 (2012) 2012;2012:228 [FREE Full text] [Medline: 24199091]
53. Brownstein JS, Huston JE, Steingold L, Joyce MV. On-demand delivery of influenza vaccination. Ann Intern Med 2015 Dec 17;163(10):806-807. [doi: 10.7326/L15-5169] [Medline: 26571249]
54. Nourani A, Ayatollahi H, Mirnia K. A smart phone application for the mothers of premature infants. Innov Res BioMed Eng 2019:269. [doi: 10.1016/j.irbm.2019.04.006]
55. Bednarczyk RA, Frew PM, Salmon DA, Whitney E, Omer SB. ReadyVax: a new mobile vaccine information app. Hum Vaccin Immunother 2017 May 4;13(5):1149-1154 [FREE Full text] [doi: 10.1080/21645515.2016.1263779] [Medline: 28059610]
56. Singh H, Mallaiah R, Yadav G, Verma N, Sawhney A, Brahmachari SK. iCHRCloud: web & mobile based child health imprints for smart healthcare. J Med Syst 2017 Dec 29;42(1):14. [doi: 10.1007/s10916-017-0866-5] [Medline: 29188446]

57. Wilson K, Atkinson K, Pluscauskas M, Bell C. A mobile-phone immunization record in Ontario: uptake and opportunities for improving public health. J Telemed Telecare 2014 Dec;20(8):476-480. [doi: 10.1177/1357633X14537771] [Medline: 25084770]
58. Wilson K, Atkinson KM, Penney G. Development and release of a national immunization app for Canada (ImmunizeCA). Vaccine 2015 Mar 30;33(14):1629-1632. [doi: 10.1016/j.vaccine.2015.02.022] [Medline: 25704801]
59. Burgess K, Atkinson KM, Westeinde J, Crowcroft N, Deeks SL, Wilson K, Public Health Agency of Canada/Canadian Institutes of Health and Research Influenza Research Network (PCIRN) Program Delivery and Evaluation Group. Barriers and facilitators to the use of an immunization application: a qualitative study supplemented with Google analytics data. J Public Health (Oxf) 2017 Sep 1;39(3):e118-e126 [FREE Full text] [doi: 10.1093/pubmed/fdw032] [Medline: 27247122]
60. Karanth SS, Lairson DR, Savas LS, Vernon SW, Fernández ME. The cost of developing a computerized tailored interactive multimedia intervention vs a print based Photonovella intervention for HPV vaccine education. Eval Program Plann 2017 Aug;63:1-6. [doi: 10.1016/j.evalprogplan.2017.02.009] [Medline: 28279866]
61. Ruiz-López T, Sen S, Jakobsen E, Tropé A, Castle PE, Hansen BT, et al. FightHPV: design and evaluation of a mobile game to raise awareness about human papillomavirus and nudge people to take action against cervical cancer. JMIR Serious Games 2019 May 8;7(2):e8540 [FREE Full text] [doi: 10.2196/games.8540] [Medline: 30958271]
62. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
63. 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 Mar 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
64. Wilson K, Mills EJ, Norman G, Tomlinson G. Changing attitudes towards polio vaccination: a randomized trial of an evidence-based presentation versus a presentation from a polio survivor. Vaccine 2005 May 27;23(23):3010-3015. [doi: 10.1016/j.vaccine.2004.12.002] [Medline: 15811647]
65. Nyhan B, Reifler J, Richey S, Freed GL. Effective messages in vaccine promotion: a randomized trial. Pediatrics 2014 May;133(4):e835-e842. [doi: 10.1542/peds.2013-2365] [Medline: 24590751]
66. Damnjanović K, Graeber J, Ilić S, Lam WY, Lep Z, Morales S, et al. Parental decision-making on childhood vaccination. Front Psychol 2018;9:735 [FREE Full text] [doi: 10.3389/fpsyg.2018.00735] [Medline: 29951010]
67. Lindley MC, Boyer-Chu L, Fishbein DB, Kolasa M, Middleman AB, Wilson T, Working Group on the Role of Schools in Delivery of Adolescent Vaccinations. The role of schools in strengthening delivery of new adolescent vaccinations. Pediatrics 2008 Jan;121(Suppl 1):S46-S54. [doi: 10.1542/peds.2007-1115F] [Medline: 18174320]
68. O'Neill J, Newall F, Antolovich G, Lima S, Danchin M. The uptake of adolescent vaccinations through the school immunisation program in specialist schools in Victoria, Australia. Vaccine 2019 Jan 7;37(2):272-279. [doi: 10.1016/j.vaccine.2018.11.034] [Medline: 30522907]
69. Paez A. Grey literature: an important resource in systematic reviews. J Evid Based Med 2017 Dec 21:- epub ahead of print. [doi: 10.1111/jebm.12265] [Medline: 29266844]


## Abbreviations

**AXIS:** Appraisal tool for Cross-Sectional Studies
**CENTRAL:** Cochrane Central Register of Controlled Trials
**EPI:** Expanded Program on Immunization
**ERIC:** Education Resources Information Center
**HPV:** human papillomavirus
**MeSH:** Medical Subject Headings
**mHealth:** mobile health
**ROBINS-I:** risk of bias in nonrandomized studies of interventions


*Edited by G Eysenbach; submitted 10.12.19; peer-reviewed by C Fernández, M Das; comments to author 03.01.20; revised version received 13.03.20; accepted 22.03.20; published 18.05.20.*

*Please cite as:*
*de Cock C, van Velthoven M, Milne-Ives M, Mooney M, Meinert E*
*Use of Apps to Promote Childhood Vaccination: Systematic Review*
*JMIR Mhealth Uhealth 2020;8(5):e17371*
*URL: https://mhealth.jmir.org/2020/5/e17371*
*doi:10.2196/17371*
*PMID:32421684*

Original Paper

# Use of the Healthy Lifestyle Coaching Chatbot App to Promote Stair-Climbing Habits Among Office Workers: Exploratory Randomized Controlled Trial

Meihua Piao[1], RN, PhD; Hyeongju Ryu[2], RN, BSN[‡]; Hyeongsuk Lee[3], RN, PhD; Jeongeun Kim[2,4,5], RN, INS, PhD

[1]Office of Hospital Information, Seoul National University Hospital, Seoul, Republic of Korea

[2]College of Nursing, Seoul National University, Jongno-gu, Republic of Korea

[3]College of Nursing, Gachon University, Incheon, Republic of Korea

[4]Interdisciplinary Program of Medical Informatics, Seoul National University, Seoul, Republic of Korea

[5]Research Institute of Nursing Science, Seoul National University, Seoul, Republic of Korea

[‡]Consumer Health Informatics & Communication Lab, Seoul National University

**Corresponding Author:**
Hyeongju Ryu, RN, BSN
College of Nursing
Seoul National University
103 Daehak-ro
Jongno-gu, 03080
Republic of Korea
Phone: 82 2 740 8803
Fax: 82 2 747 3948
Email: rnyoo@snu.ac.kr

## Abstract

**Background:** Lack of time for exercise is common among office workers given their busy lives. Because of occupational restrictions and difficulty in taking time off, it is necessary to suggest effective ways for workers to exercise regularly. Sustaining lifestyle habits that increase nonexercise activity in daily life can solve the issue of lack of exercise time. Healthy Lifestyle Coaching Chatbot is a messenger app based on the habit formation model that can be used as a tool to provide a health behavior intervention that emphasizes the importance of sustainability and involvement.

**Objective:** This study aimed to assess the efficacy of the Healthy Lifestyle Coaching Chatbot intervention presented via a messenger app aimed at stair-climbing habit formation for office workers.

**Methods:** From February 1, 2018, to April 30, 2018, a total of 106 people participated in the trial after online recruitment. Participants were randomly assigned to the intervention group (n=57) or the control group (n=49). The intervention group received cues and intrinsic and extrinsic rewards for the entire 12 weeks. However, the control group did not receive intrinsic rewards for the first 4 weeks and only received all rewards as in the intervention group from the fifth to twelfth week. The Self-Report Habit Index (SRHI) of participants was evaluated every week, and the level of physical activity was measured at the beginning and end of the trial. SPSS Statistics version 21 (IBM Corp) was used for statistical analysis.

**Results:** After 4 weeks of intervention without providing the intrinsic rewards in the control group, the change in SRHI scores was 13.54 (SD 14.99) in the intervention group and 6.42 (SD 9.42) in the control group, indicating a significant difference between the groups ($P$=.04). When all rewards were given to both groups, from the fifth to twelfth week, the change in SRHI scores of the intervention and control groups was comparable at 12.08 (SD 10.87) and 15.88 (SD 13.29), respectively ($P$=.21). However, the level of physical activity showed a significant difference between the groups after 12 weeks of intervention ($P$=.045).

**Conclusions:** This study provides evidence that intrinsic rewards are important to enhance the sustainability and effectiveness of an intervention. The Healthy Lifestyle Coaching Chatbot program can be a cost-effective method for healthy habit formation.

**Trial Registration:** Clinical Research Information Service KCT0004009; https://tinyurl.com/w4oo7md

***(JMIR Mhealth Uhealth 2020;8(5):e15085)*** doi:10.2196/15085

**KEYWORDS**

exercise; habits; reward; health behavior; healthy lifestyle

## Introduction

In a busy working life, lack of time for exercise is common among office workers. Given their occupational restrictions and difficulty in taking time off work, it is necessary to suggest effective ways for workers to exercise regularly. Nonexercise activities that can easily be performed in daily life have been introduced as an effective regimen. Nonexercise activity refers to the expending of energy via lifestyle physical activities, such as walking and stair climbing, which are naturally performed activities rather than intentional, planned, and structured activities [1]. Sustaining lifestyle habits that increase nonexercise activity in daily life can prevent weight gain even if effective weight loss cannot be expected. These physical activities are also simple behaviors that can be done habitually because they can be unconsciously repeated in daily life.

Until now, many models and theories have been proposed to explain healthy behavior in behavior-changing programs. However, these attempts are limited to cognitive areas and lack explanations for the sustainability of long-term solutions [2].

A habit formation model in which the habit is established by inducing repeated behaviors through cue-behavior-reward links in a consistent context has been suggested, and experiments are being conducted to support the effectiveness [3-5]. Habits require minimal deliberation or planning and can be enacted without conscious intention, and the key element for habit acquisition is context-specific repetition. This involves carrying out the target behavior repeatedly in the same situation to reinforce associations between the behavior and the situational cues [6]. Several treatment modalities have been designed and implemented using the habit formation model and have suggested a promising new tool to support measurable change in the way patients relate to food and physical activity. However, most of these have substantial barriers that undermine long-term strategies such as lack of adherence to the intervention, time constraints, and lack of consistent follow-ups over the long term [7].

The use of information and communication technologies, especially mobile apps, is demonstrating great potential in the delivery of treatment programs. These interventions are becoming highly valuable by promoting the continuous access of patients without the need for face-to-face meetings, home visits, or extra expenses [8]. However, most health-related apps have functions that provide information and tracking and record activity status, making continuous use difficult. The effectiveness tends to be low, with many users not using the app after downloading it and commonly deleting it after 1 month [9]. One of the reasons for this is lack of continuous motivation during management. Behaviors that require habit formation also require continuous motivation, and it is necessary to build an environment in which they form organic relationships with each other; thus, implementation of these functions in smartphone apps is necessary to provide an intervention.

In South Korea, the most popular messenger app, KakaoTalk, had 48.2 million active users as of the first quarter of 2015. KakaoTalk uses the smartphone network to deliver real-time communication in one-to-one or group chats and can be linked to services such as Kakao Plus Friend [10]. Although messenger apps are limited in functional aspects when compared with other apps, they supplement their limitations through habituation, convenience, interaction, social presence among network members, and active emotional exchange. They can therefore be used as a tool to provide a health behavior intervention that emphasizes the importance of sustainability and involvement. It is thus necessary to develop a mobile intervention delivery method that maintains interest and manages health behaviors from a long-term perspective. Thus, in this study, we aimed to assess the efficacy of the Healthy Lifestyle Coaching Chatbot intervention presented via a messenger app aimed at stair-climbing habit formation for office workers.

## Methods

### Study Design

A parallel study design was used for this exploratory trial that took place for 12 weeks from February to April 2018. Randomization was used to avoid contamination between the intervention and control groups. The research plan was reviewed and approved by the institutional review board of Seoul National University (IRB No. 1706/003-026), and the clinical trial was registered with the Clinical Research Information Service [KCT0004009].

### Recruitment

The sample size was calculated based on the power analysis formula [11]. With a significance level $\alpha$=.05, power 1–$\beta$=0.80, and medium effect size=0.5 set based on a 1-tailed test, the minimum number needed for each group was 51. After assuming a 20% dropout rate, we calculated that each group needed 61 participants for a total of 122 research participants.

Participants were employees whose work was performed in an office or another administrative setting. We recruited those who understood the purpose of the research, wanted to participate voluntarily, and gave written consent. The inclusion criteria were those aged 24 years and older who could understand and respond to the survey content and had experience using the KakaoTalk app.

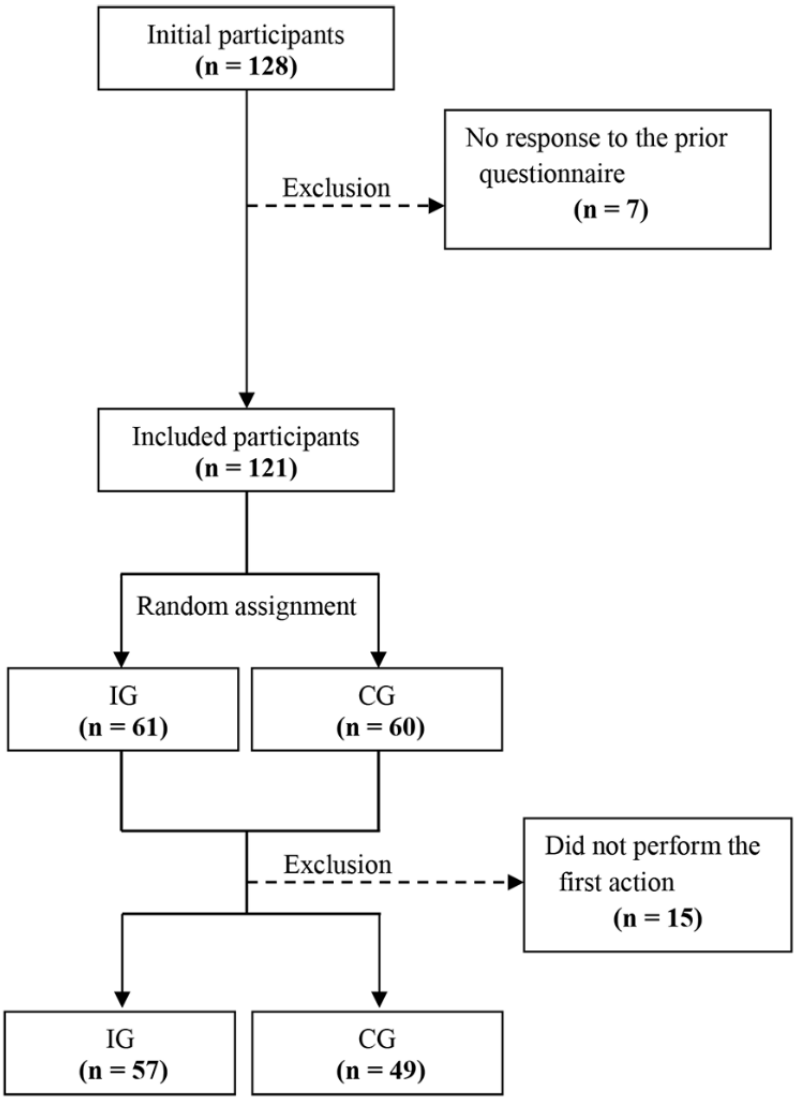
Participants were recruited from one of the advertising sharing platforms. With this platform, office workers could get information on companies, restaurants, events, and various community content. The recruitment details were uploaded on the platform event page. Once participants showed interest and agreed to join the study, they could connect to the KakaoTalk Plus Friend ID Healthy Lifestyle Coaching Chatbot through a QR code and click the “add friend” icon to be added as members automatically. An online consent form and survey link were then sent. Because only one KakaoTalk ID can be set per mobile

phone number, it is possible to check the participants from manager account in real time.

During recruitment, 128 participants who read the research introduction signed up through QR code. After 7 who did not answer the questionnaire were excluded, 121 participants were randomized into 2 groups: 61 in the intervention group (IG) and 60 in the control group (CG). Block randomizations were performed using the R program (R Foundation for Statistical Computing). The number of subjects was set as 121, and the block size was set as 4.

Because performing the first action is important in habit formation, participants were asked to perform the stair-climbing and upload a proof shot. A total of 15 participants (4 in the IG and 11 in the CG) who did not upload a picture were considered nonexecuters of the first action and were excluded from the study. A total of 106 participants (57 in the IG and 49 in the CG) were included in the 12-week intervention program (Figure 1). The intervention was performed from February 1, 2018 to April 30, 2018.

**Figure 1.** Flow of study participant enrollment.




## Messenger App-Based Intervention

The intervention Healthy Lifestyle Coaching Chatbot was presented via an app aimed at progressively establishing stair-climbing habits and increasing physical activity levels. The most popular messenger app in South Korea, KakaoTalk, was used as a platform for delivering the intervention. The functions were designed based on the habit formation model, including three elements: cue, behavior, and rewards (internal and external rewards). In addition, an automatically responding chatbot was developed using the Watson Conversation tool (IBM Corp), and it was linked to the KakaoTalk Smart Chat application programming interface (API) through the RESTful API.

In this study, cue was designed as a push alarm sent automatically to the participants every day to remind them to perform the stair-climbing behavior. Participants were asked to set realistic behavioral goals based on their daily routines, and their responses were designed as a push alarm. Through the KakaoTalk administrator messaging system, they can be individually set and automatically sent to a designated person at a specified time. The push alarms were composed of five basic conditions: who, when, where, what, and how much. For

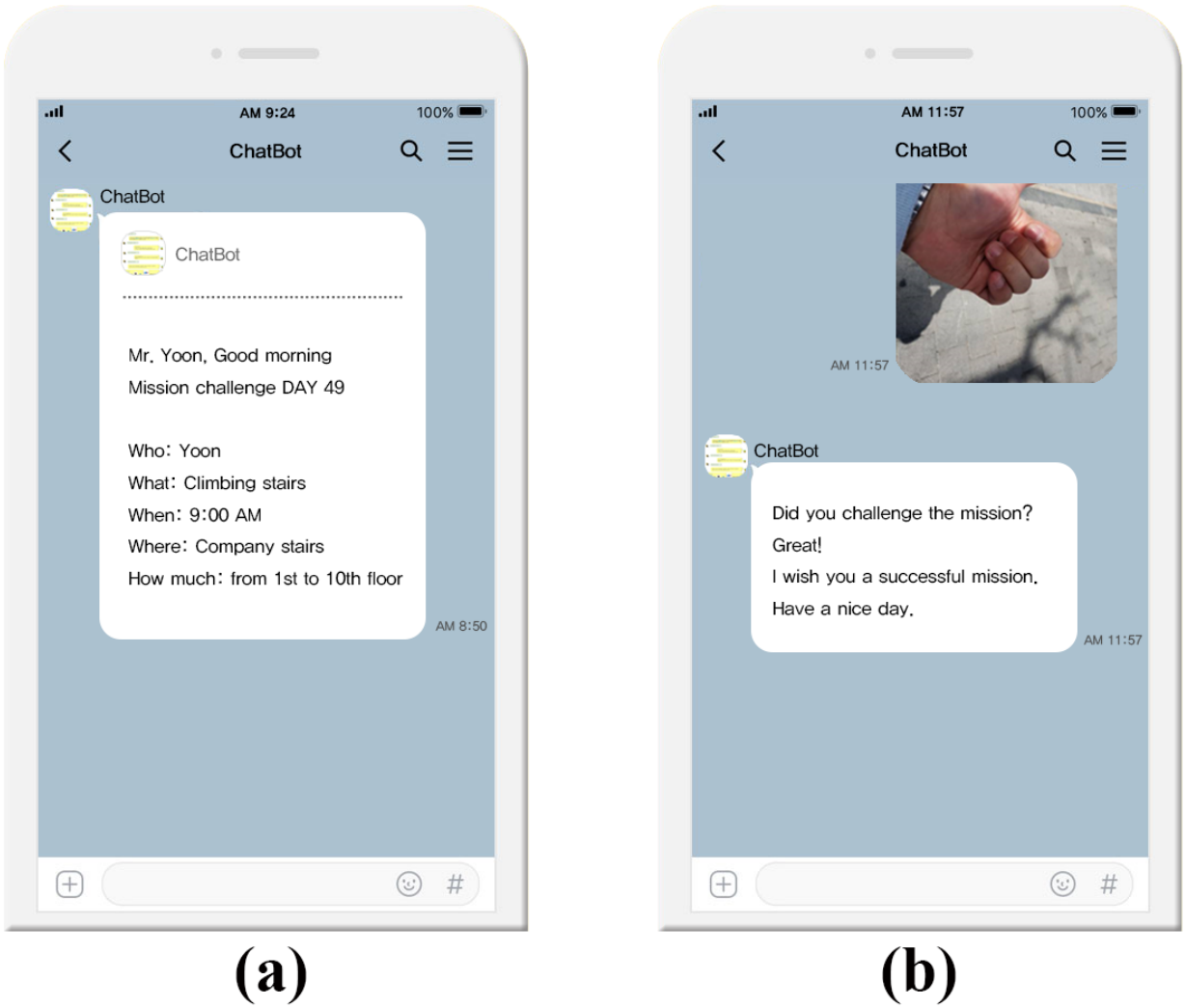
example, if a participant named John routinely arrives at the office at 9:00 am and wants to use the company building stairs to get to the fourth floor, the behavior plan should be set as who: John, when: 9:00 am, where: company building stairs, what: stair climbing, how much: fourth floor. The daily push alarm would then be sent at 8:50 am, 10 minutes before the planned time, and include the information above as a cue. Push alarms were designed to be sent from Monday through Friday, official work days. Rewards were considered reinforcers of cue-response associations. A fundamental distinction can be drawn between extrinsic rewards (eg, financial incentives) and intrinsic rewards (eg, pleasure, satisfaction) [12].

The chat scenario was designed and implemented through the Watson conversation launch tool to enable the automatic chat function. It was then linked to the app through API. Because it was linked to the Plus Friend Smart Chat API, participants could chat with the chatbot through the KakaoTalk app to receive feedback based on the behavior performed each day.

Based on the literature, we designed the scenario with two conditions. For extrinsic rewards, points and coffee coupons were provided. To encourage positive emotions, coffee coupons were sent when participants completed the first day goal-directed action. Also, a message was automatically sent at 8:00 pm every day. The participants interacted with the chatbot by chatting about the contents of their performance. If the planned behavior was completed, participants were provided with 50 points per day, and if they repeated this behavior more than 3 times a week, a coffee coupon was sent on Sunday.

For intrinsic rewards, accomplishment and positive reinforcement were used. Participants were asked to take pictures once they finished the action and upload them to the chat room as evidence of the action, after which the chatbot would automatically send a compliment message providing positive feedback. Moreover, the pictures posted by other participants were collected, edited, and sent to participants every day to acknowledge that many participants joined together to challenge the mission and receive positive reinforcement. The intervention differed between the groups. Cues and intrinsic and extrinsic rewards were provided to the IG for the entire 12 weeks. However, intrinsic rewards were excluded for the first 4 weeks in the CG and added from the fifth to twelfth weeks (Figure 2).

**Figure 2.** Screenshots of (a) push alarm reminders and (b) uploaded picture evidence and accomplishment with positive reinforcement as an intrinsic reward.



### Outcome Measures

Habit is an automatic behavior gained via repetitive process. Because it was important to measure the habit strength of a particular action in our study, we measured the outcome using the Self-Report Habit Index (SRHI) [13]. SRHI is a tool that quantitatively measures the habit strength of a particular action and is composed of 12 items. Each item is measured on a 7-point Likert scale from 1=strongly disagree to 7=strongly agree, and the total score ranges from 7 to 84 points. Cronbach α reliability of this tool was .89 [14]. The survey for measuring SRHI was designed using Google Forms, and the link was sent to the participants in the messenger app every Sunday for 12 weeks. Participants could click the link and be taken to the survey pages.

### Statistical Analysis

SPSS Statistics version 21 (IBM Corp) was used for statistical analysis. The general characteristics of the participants were analyzed using frequency, percentage, mean, and standard deviation. Relevant statistical analyses were first performed to verify proper randomization (independent samples *t* test, chi-square test, and Fisher independent sample test). Repeated measures analysis of variance (ANOVA) with 12 moments was applied for measuring the changes in SRHI scores to evaluate the effect of the intervention on the two groups. Changes in physical activity levels between groups were tested with Fisher exact tests.

## Results

### Baseline Characteristics

Testing for homogeneity of general characteristics such as sex, age, physical activity status, weight control experience, weight, and hours sitting weekly between the IG and CG showed no significant differences in any of the baseline characteristics between the two groups (Table 1).

**Table 1.** Homogeneity test of general characteristics at baseline.

| Characteristics | Intervention group (n=57) | Control group (n=49) | *t* test or chi-square | *P* value |
|---|---|---|---|---|
| **Sex, n (%)** | | | 0.001 | .92 |
| Male | 25 (44) | 21 (43) | | |
| Female | 32 (56) | 28 (57) | | |
| **Age, n (%)** | | | 5.06 | .28 |
| 20-29 | 9 (16) | 2 (4) | | |
| 30-39 | 26 (46) | 29 (59) | | |
| 40-49 | 16 (28) | 12 (25) | | |
| 50-59 | 6 (11) | 5 (12) | | |
| **Physical activity status, n (%)** | | | 0.72 | .70 |
| Low | 22 (42) | 20 (50) | | |
| Medium | 28 (54) | 18 (45) | | |
| High | 2 (4) | 2 (5) | | |
| **Weight control experience, n (%)** | | | 0.05 | .82 |
| Yes | 36 (63) | 32 (65) | | |
| No | 21 (37) | 17 (35) | | |
| Weight (kg), mean (SD) | 64.15 (14.07) | 64.47 (14.13) | –0.12 | .91 |
| Sitting hours weekly, mean (SD) | 7.55 (2.21) | 8.11 (2.58) | –1.17 | .25 |

### Attrition Rate of Participants

In the IG, 2 participants out of 57 missed the follow-up due to overseas travel in the third week. In the CG, 11 participants out of 49 were considered to have dropped out of the study. Two missed the follow-up in the first week, 3 in the second week, 5 in the third week and 1 in the fourth week. Of them all, one reported the reason for terminating participation to be overseas travel, 1 reported hospitalization, and other 9 participants were automatically terminated due to consecutive weeks of nonresponse.

### Changes of Habit Strength

Over 12 weeks, the SRHI scores increased by an average of 31.38 points in the IG and 21.04 points in the CG on the 84-point scale (Table 2). Additionally, analysis of the change in SRHI scores for each group using repeated measures ANOVA showed statistically significant changes in the scores for both groups with an increase in the duration of the intervention ($P<.001$). In addition, the SRHI scores had significant between-group differences in the two groups ($P=.008$; Table 3 and Figure 3).

**Table 2.** Self-Report Habit Index characteristics based on group.

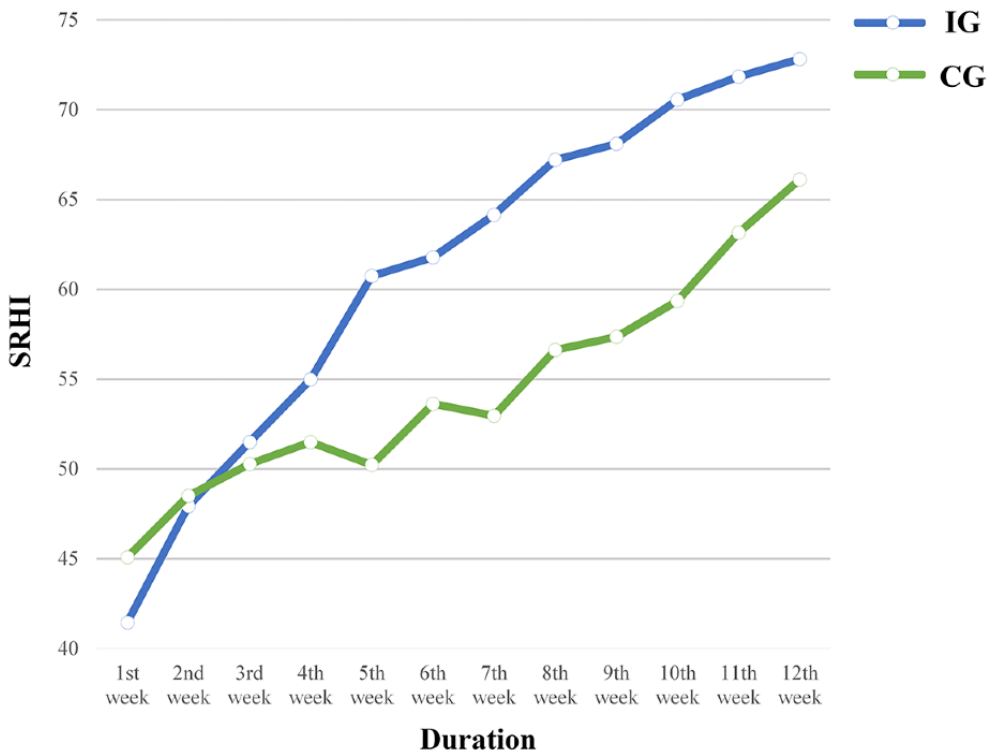
| Intervention week | Intervention group, mean (SD) | Control group, mean (SD) |
|---|---|---|
| 1 | 41.44 (15.69) | 45.08 (14.39) |
| 2 | 47.92 (16.40) | 48.50 (13.66) |
| 3 | 51.49 (16.47) | 50.27 (13.66) |
| 4 | 54..97 (16.92) | 51.50 (13.60) |
| 5 | 60.74 (14.62) | 50.23 (11.96) |
| 6 | 61.79 (14.53) | 53.62 (13.33) |
| 7 | 64.15 (14.37) | 52.96 (16.26) |
| 8 | 67.21 (14.80) | 56.62 (15.37) |
| 9 | 68.10 (11.64) | 57.35 (10.39) |
| 10 | 70.56 (10.06) | 59.35 (10.39) |
| 11 | 71.85 (9.71) | 63.15 (10.68) |
| 12 | 72.82 (9.77) | 66.12 (7.15) |

**Table 3.** Multivariate test results.

| Effect | Value | *F* test | Hypothesis df[a] | Error df | Significance |
|---|---|---|---|---|---|
| **Habit** | | | | | |
| Pillai's trace | 0.815 | 21.16 | 11 | 75 | <.001 |
| Wilks' lambda | 0.185 | 21.16 | 11 | 75 | <.001 |
| **Habit * Group** | | | | | |
| Pillai's trace | 0.359 | 2.70 | 11 | 75 | .008 |
| Wilks' lambda | 0.641 | 2.70 | 11 | 75 | .008 |

[a]df: degree of freedom.

**Figure 3.** Plots of Self-Report Habit Index (SRHI) automaticity scores.



### Effects of Intrinsic Rewards

After the fourth week of the intervention, the changes in the SRHI scores between the two groups were compared. The total change in SRHI scores had increased by 13.54 (SD 14.99) points in the IG and 6.42 (SD 9.42) points in the CG, and there were significant differences between the two groups during that period (*P*=.04).

After the fifth week, the intrinsic reward was added to the intervention in the CG, and the same intervention was provided to both groups from the fifth to twelfth week. We found that the total change in SRHI scores was 12.08 (SD 10.87) points in the IG and 15.88 (SD 13.29) points in the CG, with no significant difference in the changes between the two groups from the fifth to the twelfth week (*P*=.21; Table 4). This result indicated significant differences in SHRI score changes between the two groups for the first 4 weeks where no intrinsic reward was provided to the CG. However, all scores increased with no significant difference between the groups as a result of applying identical interventions to both groups by adding the intrinsic rewards to the CG from the fifth week. This indicated that the intrinsic reward acted as an influencing factor in habit formation.

**Table 4.** Self-Report Habit Index variation based on the difference of rewards.

| Intervention duration | Variation | Intervention group, mean (SD) | Control group, mean (SD) | t(*p*) |
|---|---|---|---|---|
| Weeks 1 to 4 | ΔSRHI[a] (4–1) | 13.54 (14.99) | 6.42 (9.42) | 2.12 (.04) |
| Weeks 5 to 12 | ΔSRHI (12–5) | 12.08 (10.87) | 15.88 (13.29) | 0.21 (.21) |

[a]ΔSRHI: change in Self-Report Habit Index.

## Discussion

### Principal Findings

This study shows that the Healthy Lifestyle Coaching Chatbot program resulted in effective habit formation for stair-climbing behavior. Forming a habit is challenging, especially when transforming many things at once. In this study, the participants were asked to set small goals and take things one step at a time. Also, the behavior must be carried out repeatedly in the presence of the same contextual cues. Habit is an automatic behavior gained via repetition. In this study, most participants tended to increase the number of repetitions of the behavior and thus increase the habit strength as the intervention proceeded. This suggests that, according to the habit formation model hypothesis, given a consistent context, the repeated occurrence of cue-behavior-reward chains leads to increases in the strength of the habit behavior [15].

Based on the literature, a fundamental distinction can be drawn between extrinsic rewards (eg, financial incentives) and intrinsic rewards (eg, pleasure, satisfaction) [16]. The reward acts as motivation for sustaining a new behavior, so is an important factor in the repeated execution of behavior. However, previous research has suggested that providing simple financial rewards has limited success in sustaining motivational inducement. Although extrinsic rewards elicit motivation in the beginning, they eventually hinder a particular behavior from becoming habitual [16]. In this study, the SRHI score was 3.64 points higher in the CG after the first week oof the intervention. However, the upward trend until the fourth week was slower than the IG when only extrinsic reward was applied. Analysis of the SRHI score changes after the fourth week of the intervention showed that the changes were greater for the IG than for the CG at a statistically significant level. Furthermore, when similar rewards (including intrinsic and extrinsic) were provided to both groups from the fifth week on, the SRHI score showed a similar slope and increase without a significant between-group difference. This result concurs with other studies that found that extrinsic rewards elicit motivation regarding the behavior initially, but later the intrinsic rewards played an important factor for increasing the behavior continuation [17].

Lally et al [5] tracked the formation of healthy habits in a naturalistic setting, repeating a behavior in the presence of consistent cues. Based on the results, the habit formation tracked using SRHI was found to typically follow an asymptotic curve, and the initial repetitions caused large increases in SRHI scores, but with each new repetition, the score gains reduced until the behavior reached its limit of automaticity. Also, the average time for participants to reach the asymptote of automaticity was 65 days, which is near 10 weeks. In this study, changes in SRHI scores showed lower for the IG compared to the CG after week 5 even though there was no significant difference between the groups. We found that for the IG, the SRHI score increased by an average of 1 point after the tenth week. This result showed that it was possible for a repeated behavior to reach its maximum level of automaticity for the IG during the 12-week intervention. However, for the CG, the SRHI score increased by an average of 3 points after the tenth week. It is interesting to note that an intrinsic reward could help participants perform the behavior consistently enough to achieve habit status. The results of our study suggest the need for interventions using intrinsic reward factors to design an intervention program. Furthermore, positive feedback such as a compliment as an intrinsic reward is a simple but powerful tool in which the positive value of the reward can act as a motivational factor [16].

In this study, the messenger app KakaoTalk was used as an intervention delivery method. The KakaoTalk app uses the smartphone network to deliver real-time communication in one-to-one or group chats and can be linked to services such as chatbots. KakaoTalk has superior habituation, convenience, and interaction and is a platform in which social presence among network members and emotional exchanges are actively occurring [10]. With advancements in digital technology, mobile apps that help exercise management are being released, and wearable devices are becoming increasingly diverse. However, the dropout rate was considered to be high. One of the causes was continuous motivation that occurs during management.

In this study, 11 participants out of 49 in the CG and 2 participants out of 57 in the IG dropped out during the intervention; the total dropout rate was 12.26%. Also, all dropout cases appeared during the first 4 weeks. In comparison with previous studies in which dropout rates approached 50% at week 4, our study showed comparatively lower dropout rates. Although mobile interventions offer a promising way to deliver related content, dropout rates for this form may be high [18,19]. One potential contributing factor for this dropout rate is the difficulty in the use of the app, particularly when the app is designed to be used in a self-guided fashion [20].

This study designed the method of intervention delivery using a familiar app. Participants did not need to download other apps and learn how to use them. This delivery method is significant in that it minimized participant aversion, can provide sustained motivation, and can form an organic relationship.

For the duration of the intervention period, most participants dropped out in week 3 and week 4 during which the Lunar New Year holidays occurred. Based on prior research, lapses have an important influence on the sustainability of behavior. Although lapses of approximately 1 to 2 days do not greatly affect the sustainability of behavior, lapses of approximately 1 week greatly affect the sustainability of behavior and significantly lower the possibility of future execution of behavior [5]. We found that research subjects who lapsed for approximately 1 week due to the Lunar New Year on week 3 also showed high dropout rates in our study as well. This suggests the need to improve the effectiveness of the program by including intervention methods to provide motivation so that lapses do not exceed 1 week by specially managing the research subjects who are lapsing.

Thus, using a messenger app that allows a continuous relationship with the research subjects and provides motivation and professional advice is garnering much interest. Therefore, the methodology used in our study can contribute to enhancing the sustainability and effectiveness of care intervention by developing programs and delivery tools that implement theories on a wider variety of illnesses or health behaviors.

### Limitations

There were limitations in designing the program because the functionality provided by the current messenger app does not include direct communication functions, such as informing colleagues of their commitments and receiving positive support messages from others. Also, research should be done to implement different kinds of rewards to identify how the rewards affect the habits and which ones are the most important to maintain habits. Furthermore, there could be possible bias in the results since some studies have shown that participants who volunteer in these kinds of research are more motivated to be physically active or develop the habits.

### Conclusions

This study provides evidence that the Healthy Lifestyle Coaching Chatbot program presented via an app can be used to provide improvements in relevant variables in the long term and can be useful in producing new healthy habits. This messenger app as a delivery tool is considered a cost-effective means to deliver interventions to a large number of people.

### Acknowledgments

This paper is based on parts of MP's doctoral dissertation.

### Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2403 KB - mhealth_v8i5e15085_app1.pdf ]

### References

1. Levine JA. Non-exercise activity thermogenesis (NEAT). Best Pract Res Clin Endocrinol Metab 2002 Dec;16(4):679-702. [doi: 10.1053/beem.2002.0227] [Medline: 12468415]
2. Salema NM, Elliott RA, Glazebrook C. A systematic review of adherence-enhancing interventions in adolescents taking long-term medicines. J Adolesc Health 2011 Nov;49(5):455-466. [doi: 10.1016/j.jadohealth.2011.02.010] [Medline: 22018559]
3. Lally P, Gardner B. Promoting habit formation. Health Psychol Rev 2013 May;7(sup1):S137-S158. [doi: 10.1080/17437199.2011.603640]
4. Wood W, Rünger D. Psychology of habit. Annu Rev Psychol 2016 Jan;67:289-314. [doi: 10.1146/annurev-psych-122414-033417] [Medline: 26361052]
5. Lally P, van Jaarsveld CHM, Potts HWW, Wardle J. How are habits formed: modelling habit formation in the real world. Eur J Soc Psychol 2009 Jul 16;40(6):998-1009. [doi: 10.1002/ejsp.674]
6. McGowan L, Cooke LJ, Gardner B, Beeken RJ, Croker H, Wardle J. Healthy feeding habits: efficacy results from a cluster-randomized, controlled exploratory trial of a novel, habit-based intervention with parents. Am J Clin Nutr 2013 Sep;98(3):769-777. [doi: 10.3945/ajcn.112.052159] [Medline: 23864536]

7. Mensorio MS, Cebolla-Martí A, Rodilla E, Palomar G, Lisón JF, Botella C, et al. Analysis of the efficacy of an internet-based self-administered intervention ("Living Better") to promote healthy habits in a population with obesity and hypertension: an exploratory randomized controlled trial. Int J Med Inform 2019 Apr;124:13-23. [doi: 10.1016/j.ijmedinf.2018.12.007] [Medline: 30784422]
8. Nilsen W, Kumar S, Shar A, Varoquiers C, Wiley T, Riley WT, et al. Advancing the science of mHealth. J Health Commun 2012 May;17 Suppl 1:5-10. [doi: 10.1080/10810730.2012.677394] [Medline: 22548593]
9. Do Y, Keum S, Lee S, Lee J. The exploratory study of factors which influence to the maintenance of using wearable device for healthcare: interdisciplinary approach to user experiences combined with technology, psychology and interaction perspectives. Comm Korean Inform Sci Soc 2014;32(11):37-45.
10. Choi MJ, Kim WG. An exploratory study on use motivation, evaluation and selection of SNS mobile marketing services focused on the "KakaoTalk PlusChingu". Korean J Advert Public 2013;15(2):85-110.
11. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
12. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
13. Orbell S, Verplanken B. The automatic component of habit in health behavior: habit as cue-contingent automaticity. Health Psychol 2010 Jul;29(4):374-383. [doi: 10.1037/a0019596] [Medline: 20658824]
14. Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Social Pyschol 2003 Jun;33(6):1313-1330. [doi: 10.1111/j.1559-1816.2003.tb01951.x]
15. Gardner B, de Bruijn G, Lally P. A systematic review and meta-analysis of applications of the Self-Report Habit Index to nutrition and physical activity behaviours. Ann Behav Med 2011 Oct;42(2):174-187. [doi: 10.1007/s12160-011-9282-0] [Medline: 21626256]
16. 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):627-668. [doi: 10.1037/0033-2909.125.6.627] [Medline: 10589297]
17. Verplanken B. Beyond frequency: habit as mental construct. Br J Soc Psychol 2006 Sep;45(Pt 3):639-656. [doi: 10.1348/014466605X49122] [Medline: 16984725]
18. Armitage CJ. Can the theory of planned behavior predict the maintenance of physical activity? Health Psychol 2005 May;24(3):235-245. [doi: 10.1037/0278-6133.24.3.235] [Medline: 15898858]
19. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
20. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]


## Abbreviations

**ANOVA:** analysis of variance
**API:** application programming interface
**CG:** control group
**IG:** intervention group
**SRHI:** Self-Report Habit Index


*Edited by G Eysenbach; submitted 15.07.19; peer-reviewed by L Grepo, H Santos, V Manea; comments to author 03.10.19; revised version received 13.11.19; accepted 07.02.20; published 19.05.20.*

*Please cite as:*
*Piao M, Ryu H, Lee H, Kim J*
*Use of the Healthy Lifestyle Coaching Chatbot App to Promote Stair-Climbing Habits Among Office Workers: Exploratory Randomized Controlled Trial*
*JMIR Mhealth Uhealth 2020;8(5):e15085*
*URL: https://mhealth.jmir.org/2020/5/e15085*
*doi:10.2196/15085*
*PMID:32427114*

Original Paper

# A Pedometer-Guided Physical Activity Intervention for Obese Pregnant Women (the Fit MUM Study): Randomized Feasibility Study

Jai N Darvall[1,2], MBBS, MEpid, FANZCA, FCICM; Andrew Wang[3], MD; Mohamed Nusry Nazeem[3], MD; Cheryce L Harrison[4], PhD, AEP; Lauren Clarke[5], BPT; Chennelle Mendoza[5], BPT; Anna Parker[2], BA (Hons), MA, MBioethics; Benjamin Harrap[6], BA (Hons), MBiostat; Glyn Teale[7], MRCP, MD, FRANZCOG; David Story[2], BMedSc, MBBS, MD, FANZCA; Elizabeth Hessian[8], MBBS, FANZCA

[1]Department of Anaesthesia and Pain Management, Royal Melbourne Hospital, Melbourne, Australia

[2]Centre for Integrated Critical Care, University of Melbourne, Melbourne, Australia

[3]Melbourne Medical School, University of Melbourne, Melbourne, Australia

[4]Monash Centre for Health Research and Implementation, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia

[5]Department of Physiotherapy, Western Health, Melbourne, Australia

[6]Melbourne Epicentre, University of Melbourne, Melbourne, Australia

[7]Department of Women's and Children's Services, Western Health, Melbourne, Australia

[8]Department of Anaesthesia and Pain Management, Western Health, Melbourne, Australia

**Corresponding Author:**
Jai N Darvall, MBBS, MEpid, FANZCA, FCICM
Department of Anaesthesia and Pain Management
Royal Melbourne Hospital
300 Grattan St, Parkville
Melbourne, 3050
Australia
Phone: 61 393427000
Email: jai.darvall@mh.org.au

## Abstract

**Background:** Obesity in pregnancy is a growing problem worldwide, with excessive gestational weight gain (GWG) occurring in the majority of pregnancies. This significantly increases risks to both mother and child. A major contributor to both prepregnancy obesity and excessive GWG is physical inactivity; however, past interventions targeting maternal weight gain and activity levels during the antenatal period have been ineffective in women who are already overweight. Pedometer-guided activity may offer a novel solution for increasing activity levels in this population.

**Objective:** This initial feasibility randomized controlled trial aimed to test a pedometer-based intervention to increase activity and reduce excessive GWG in pregnant women.

**Methods:** We supplied 30 pregnant women with obesity a Fitbit Zip pedometer and randomized them into 1 of 3 groups: control (pedometer only), app (pedometer synced to patients' personal smartphone, with self-monitoring of activity), or app-coach (addition of a health coach–delivered behavioral change program). Feasibility outcomes included participant compliance with wearing pedometers (days with missing pedometer data), data syncing, and data integrity. Activity outcomes (step counts and active minutes) were analyzed using linear mixed models and generalized estimating equations.

**Results:** A total of 30 participants were recruited within a 10-week period, with a dropout rate of 10% (3/30; 2 withdrawals and 1 stillbirth); 27 participants thus completed the study. Mean BMI in all groups was ≥35 kg/m$^2$. Mean (SD) percentage of missing data days were 23.4% (20.6%), 39.5% (32.4%), and 21.1% (16.0%) in control, app group, and app-coach group patients, respectively. Estimated mean baseline activity levels were 14.5 active min/day and 5455 steps/day, with no significant differences found in activity levels between groups, with mean daily step counts in all groups remaining in the sedentary (5000 steps/day) or low activity (5000-7499 steps/day) categories for the entire study duration. There was a mean decrease of 7.8 steps/day for each increase in gestation day over the study period (95% CI 2.91 to 12.69, *P*=.002).

**Conclusions:** Activity data syncing with a personal smartphone is feasible in a cohort of pregnant women with obesity. However, our results do not support a future definitive study in its present form. Recruitment and retention rates were adequate, as was activity data syncing to participants' smartphones. A follow-up interventional trial seeking to reduce GWG and improve activity in this population must focus on improving compliance with activity data recording and behavioral interventions delivered.

**Trial Registration:** Australian and New Zealand Clinical Trials Registry ACTRN12617000038392; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=370884

***(JMIR Mhealth Uhealth 2020;8(5):e15112)*** doi:10.2196/15112

**KEYWORDS**

gestational weight gain; pregnancy; maternal obesity; lifestyle intervention; pedometer

## Introduction

Obesity in pregnancy is an endemic and growing problem worldwide. In Australia, 50% of all women who become pregnant are either overweight (BMI 25-30 kg/m$^2$) or women with obesity (BMI≥30 kg/m$^2$), corresponding with other developed countries [1-3]. Excess gestational weight gain (GWG) above US Institute of Medicine recommendations (which for women with obesity should not exceed 9 kg) occurs in the majority of pregnancies, with every kilogram above recommendations increasing adverse outcomes by 10% [4-6]. Excessive GWG in the presence of preexisting obesity exacerbates health risks for mother and child, including increased rates of gestational hypertension and diabetes, cesarean delivery, perinatal mortality, and neonatal hypoglycemia, jaundice, and admission to neonatal intensive care [1].

Although a complex problem, a significant contributor to prepregnancy obesity and excessive GWG is physical inactivity; similarly, interventions that have succeeded in increasing physical activity through exercise have been associated with a commensurate reduction in GWG in pregnant women. A recent Cochrane review encompassing over 14,000 women across 49 randomized controlled trials found a pooled reduction of 21% in excessive GWG with exercise interventions [7-9]. The evidence for increasing normal physical activity during the day, however, is less robust. World Health Organization recommendations for adults include at least 150 min of moderate-intensity, or at least 75 min of vigorous-intensity, aerobic physical activity per week, or a combination of both [10]. During pregnancy, targets for moderate-intensity physical activity are the same in the absence of contraindications [11,12]. Evidence also suggests pregnant women with obesity are less active than their pregnant normal weight counterparts; 2 recent studies using pedometers to examine activity levels in overweight and pregnant women with obesity, for example, demonstrated mean activity levels in the *sedentary* range (5000 steps/day) [13-16].

How to address this increasing problem is not clear. Frustratingly, despite the benefit of exercise in reducing GWG presented above, interventions seeking to target an increase in activity levels during the antenatal period have commonly failed in reducing GWG in the cohort of women who are already overweight [7,17]. However, there is emerging evidence that pedometer-guided activity interventions may be successful in increasing activity levels in both pregnant and nonpregnant populations [18]. The latest generation of pedometers also have the capacity to automatically upload activity information via smartphone to enable data capture and daily monitoring, which can be used to provide remote feedback to patients. Increasing smartphone ownership worldwide means there is scope for broad population reach, potentially overcoming some of the barriers to engagement with traditional models of health care, such as transport, cost, and rigidity of appointment times. What is not well understood from previous research, however, is the efficacy of smartphone app, data capturing availability, and biofeedback provided in the context of interventions to optimize GWG.

Having previously demonstrated the acceptability and utility of the Fitbit Zip pedometer as a remote activity monitoring device in a pilot study [19], we conducted a feasibility randomized controlled trial of a pedometer-based intervention in a cohort of pregnant women with obesity. In particular, we aimed to evaluate the feasibility of self-monitoring of activity levels via the Fitbit Zip pedometer and the additional role of a behavioral intervention in reducing the incidence of excessive GWG.

## Methods

This randomized, controlled feasibility trial was conducted in the Department of Anesthesia and Pain Management, Sunshine Hospital, Victoria, Australia. Approval was gained from the hospital Human Research and Ethics Committee (December 22, 2016, Human Research and Ethics Committee approval number HREC/16/MH/320), and the trial was registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12617000038392, January 10, 2017). Female patients aged ≥18 years with BMI (weight in kg/height in m$^2$) ≥30 kg/m$^2$ with the availability of a smartphone capable of allowing Fitbit data uploading (eg, Apple iPhone or Android operating system equipped phone) were eligible for enrollment between gestational week 12 and 16. A convenience sample of 30 nonconsecutive patients attending the antenatal clinic was enrolled between March and May 2017 based on the availability of study investigators, after written informed consent. Patients were excluded if they had preeclampsia, twin or multiple pregnancies, preterm rupture of membranes, incompetent cervix/cerclage, or if they had a joint or muscle disorder sufficient to impair walking to a target of 10,000 steps daily. Patients were randomized to 1 of 3 groups (2 interventional and 1 control) via a computerized random number generator, with sequentially numbered envelopes used for allocation

concealment. Because of the nature of the intervention, neither patients nor study investigators were blind to group allocation.

### Intervention

All patients were supplied with a Fitbit Zip pedometer and instructed to wear it daily on the waistband of clothing during waking hours. The pedometer measured step counts and minutes each day spent at various activity levels, characterized as either *sedentary*, *lightly*, *fairly*, or *very* active based on the cadence of steps recorded. *Active* minutes were commenced once activity exceeded 3 metabolic equivalents for ≥10 min [20]. Patients in the app and app-coach groups also had the pedometer synced to their personal smartphone via the Fitbit app (Fitbit Inc, San Francisco, California, USA), allowing automatic daily uploading of activity data. Each patient was registered on this platform under a de-identified email address. We considered days with >1000 steps reported as indicative of a day wearing the pedometer. Days with <1000 steps reported were censored as missing. Group conduct was as follows: (1) in the *control group*, the pedometer display was obscured using tamperproof tape, blinding patients to daily steps, and active minutes. To further ensure blinding, pedometers in the control group were not linked to participants' smartphones, but instead were synced manually by study investigators at clinic appointments; (2) in the *app-only intervention group (app group)*, the pedometer was synced to patients' personal smartphones, with patients encouraged to self-monitor daily step counts and activity minutes via the pedometer display or the Fitbit app; and (3) in the *app and coach intervention group (app-coach group)*, in addition to the intervention in the app group above, patients were administered a behavioral change program delivered by trained health coaches. This consisted of an initial 1-hour face-to-face session between 16 and 20 weeks of gestation at which goal setting for activity in pregnancy was discussed, including challenges and barriers to achievement, and *specific*, *measurable*, *achievable*, *relevant*, and *time-bound* [21] objectives were set. Patients then had 3 follow-up health coach 20-min telephone sessions at 24, 28, and 32 weeks of gestation, during which pedometer activity levels were reviewed, and strategies to achieve targets reinforced. This included exploring and resolving ambivalence, providing encouragement, and ensuring skills were practiced and action plans completed. The intervention was based on self-determination theory, including action planning, goal setting, and self-monitoring, all factors in long-term behavioral change [22]. The aims were to educate women about the importance of physical activity and healthy eating during pregnancy (commensurate with the guidelines below), the balance between energy intake and expenditure, and removing misconceptions and increasing confidence about engaging in physical activity throughout pregnancy (including advice on activity targets below and reassurance about the safety of vigorous exercise). In the event of patients not being contactable for follow-up telephone sessions, a total of 3 attempted phone calls were made over 2 weeks. The health coaches (CM and LC) were trained over 2 weeks in the delivery of the intervention, including motivational interviewing, by the developer of the program (CLH).

All patients, regardless of group allocation, were provided with written resources at enrollment, including the Australian Physical Activity and Sedentary Behavior Guidelines [23], the guidelines on pregnancy and exercise published by Physical Activity Australia [12], and Dietary Guidelines for Australian Adults [24]. General physical activity advice framed around these guidelines was provided at baseline—at least 30 min of moderate physical activity most days of the week, or 150 min per week. Examples given by Physical Activity Australia include brisk walking, dancing, cleaning windows or sweeping, or pushing a stroller, equating to a level of physical activity of *fairly* or *very* active as measured by the pedometer. Patients were also provided with information regarding step count categorization: a step count <5000 steps/day is classified as sedentary, 5000 to 7499 steps per day as low activity, 7500 to 10,000 steps per day as fairly active, and >10,000 steps as active [13]. This guidance was based on advice from the state tertiary obstetric referral hospital regarding a target of 10,000 steps/day [25] and evidence in obstetric populations that a step count exceeding 10,000 steps/day results in a reduction in GWG [9].

### Data Collection

Baseline demographic data collected directly from patients included age, parity, country of birth, educational background, and household income (previously shown to influence activity levels in pregnancy) [26]. Weight and height were directly measured in the antenatal clinic, and baseline BMI calculated. Activity data collected included daily step counts and daily active minutes (*fairly active* plus *very active* minutes). Absolute GWG (baseline to 36-37 weeks of gestation) was calculated from directly measured weights in the antenatal clinic.

### Outcomes

The primary aim of this feasibility trial was to refine and test the trial protocol for a follow-on large, multicenter trial. Specific feasibility outcomes were recruitment feasibility, engagement and recruitment rate, maintenance of blinding of the control group to pedometer step count (concealment of pedometer display with tamperproof tape), participant compliance with wearing pedometers (days with missing pedometer data) and syncing data regularly, participant retention to study conclusion, and data integrity and completeness of uploaded step counts to investigators. Further secondary aims to guide a definitive multicenter trial were to examine efficacy in increasing step count to a target of 10,000 steps daily in pregnant women with obesity via feedback from the pedometer, evaluate the added benefit of investigator feedback compared with participant self-monitoring alone on the reduction in excessive GWG of participants, and assess the magnitude of any effect to further inform sample size calculation for a definitive trial.

### Statistical Analysis

Data were summarized using mean (SD), median (IQR), or number (%) as appropriate. The baseline variables (age, BMI, and gestation day at recruitment) and outcome variables (steps and activity level) were first examined for the linearity of association. Step count and minutes active were analyzed using a linear mixed model with random intercepts (minutes active) and multiple linear regression using generalized estimating equations (step count), as the maximum likelihood estimation did not converge in the linear mixed model approach for step

count. Variables were centered around a gestation of 100 days, BMI of 35 kg/m$^2$, and age of 30 years. The linear mixed-model approach allowed controlling for activity level at baseline, as gestational age at enrollment varied between participants. BMI, age, parity at recruitment, and gestational period (classified into months of gestation) were also controlled for in the statistical model. Complete case analysis was used, with observations containing missing data not included in analyses. Statistical analyses were performed using Stata 14.1 (Stata Corp).

### Sample Size Calculation for Definitive Follow-Up Study

No formal sample size was calculated for this feasibility trial. A sample of 30 women was considered adequate to provide data on the feasibility outcomes listed. To test the trial protocol and feasibility endpoints, we aimed to enroll 30 patients. A follow-up definitive randomized controlled two-arm trial sample size was calculated based on the reduction in excess GWG with exercise interventions in women with overweight and obesity contained within the aforementioned Cochrane review [7]; 62% of control patients vs 52% of exercise intervention patients experienced excessive GWG, with a risk ratio of 0.84 (95% CI 0.73 to 0.95). At a power of 0.90 and an alpha error of .05, 533 patients in each group (1066 patients in total) would be required in a follow-up interventional trial.

## Results

A total of 30 patients (10 per group) were enrolled in the 3 groups (control, app, and app-coach), of whom 2 participants withdrew without activity data recorded (1 from the control group and 1 from the app group) and were subsequently excluded from the analysis. An additional control patient had a stillbirth at 29 weeks of gestation, with no further data collection. Outcome data were thus available for a total of 27 patients (Figure 1). In the app-coach group, 2 patients failed to attend their initial assessments, and no further intervention contact was made. The initial in-person intervention was delivered for the remaining 8 app-coach group patients between gestational week 16 and 20, with 4 patients continuing to completion of all 3 scheduled telephone calls. Group demographics were similar at recruitment, with mean BMI in all groups ≥35 kg/m$^2$ (Table 1). There were no differences in country of birth, educational background, and household income between groups.

**Figure 1.** Flowchart of study participants through the trial.

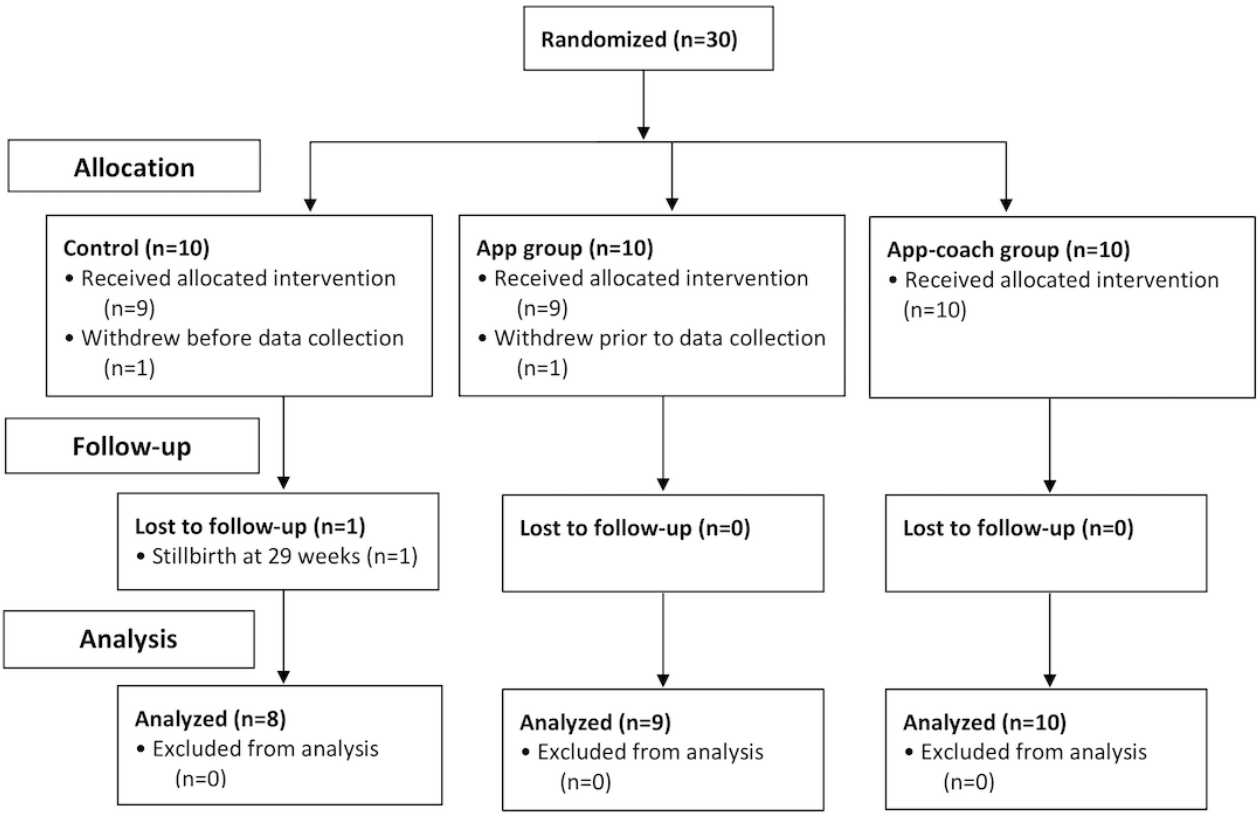


**Table 1.** Baseline characteristics (total enrolled cohort).

| Variable | Control (n=10), mean (SD) | App group (n=10), mean (SD) | App-coach group (n=10), mean (SD) |
|---|---|---|---|
| Age at recruitment (years) | 30.2 (5.3) | 30.0 (5.0) | 28.4 (5.8) |
| BMI at recruitment (kg/m$^2$) | 35.9 (4.4) | 36.7 (4.4) | 37.0 (4.2) |
| Height (cm) | 163.2 (6.0) | 166.0 (6.8) | 161.2 (8.5) |
| Weight at recruitment (kg) | 95.2 (8.4) | 101.5 (16.0) | 96.3 (15.6) |
| Gestation day at recruitment | 110 (21.0) | 105 (12.0) | 112 (19.0) |

## Feasibility Outcomes

Recruitment and retention rates were feasible, with all 30 participants recruited within a 10-week period, and a dropout rate of 10% (2 withdrawals and 1 stillbirth). Target population recruitment feasibility was also adequate, with an annual caseload of >1000 pregnant women with obesity seen at Sunshine Hospital. Control group blinding was adequate, with concealment of pedometer display maintained at each check. Patient compliance with wearing pedometers was problematic, with a percentage of days with missing data mean (SD) of 23.4% (20.6%), 39.5% (32.4%), and 21.2% (16.0%) in control, app, and app-coach groups, respectively. Over the study duration, 4 pedometers were lost, requiring replacement. Overall, regular data syncing via automatic mobile phone connection was feasible in app and app-coach group patients, although required troubleshooting in 5 women (1 manually and 4 remotely via telephone).

## Activity Data

There was no evidence of a nonlinear association between the baseline and outcome variables. Therefore, the variables were entered into the statistical models without transformation. Results of the linear mixed model investigating activity level are presented in Table 2. The estimated mean baseline daily active minutes for a 30-year-old nulliparous control patient with a BMI of 35 kg/m$^2$ and between 61 and 90 gestational days was 14.5 min. Compared with control patients, there was no difference in active minutes for patients in the app or app-coach groups. There were also no significant differences for any group in activity level trends across the gestational period (Figure 2).

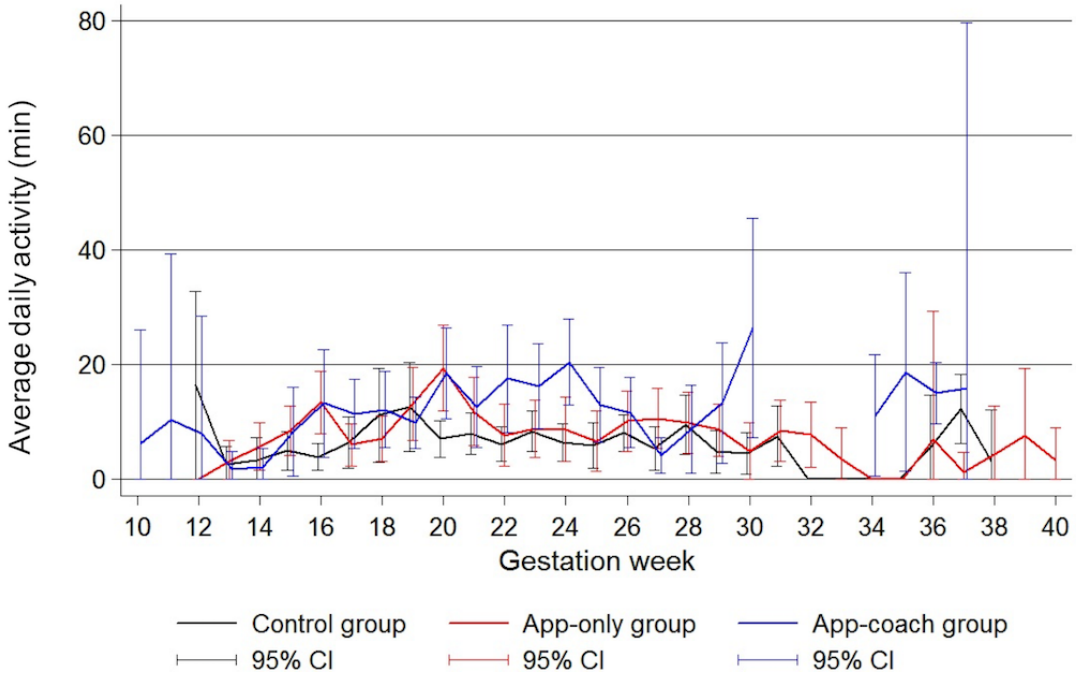
A 1-year increase in age was associated with an estimated increase in the daily activity of 0.6 min (95% CI 0.1 to 1.2 min, $P$=.03, and a 1 kg/m$^2$ increase in BMI was associated with an estimated reduction in the daily activity of 0.9 min (95% CI 0.3 to 1.5 min, $P$=.005). The estimated effect for parity was each previous live birth being associated with a decrease in the daily activity of 4.5 min (95% CI 0.7 to 8.2 minutes, $P$=.02).

**Table 2.** Fixed-effect estimates from the linear mixed model investigating activity level.

| Variable | Coefficient | Standard error | *P* value | 95% CI |
|---|---|---|---|---|
| App group | 2.11 | 3.35 | .53 | −4.46 to 8.68 |
| App-coach group | 1.82 | 3.20 | .57 | −4.45 to 8.10 |
| Gestational month | −0.49 | 0.41 | .24 | −1.30 to 0.32 |
| BMI | −0.91 | 0.32 | .005 | −1.53 to −0.28 |
| Age | 0.64 | 0.29 | .03 | 0.08 to 1.20 |
| Parity | −4.47 | 1.91 | .02 | −8.21 to −0.72 |
| Baseline[a] | 14.54 | 3.18 | reference | 8.31 to 20.76 |

[a]Baseline represents baseline activity in minutes for an individual in the control group between 61 and 90 gestational days with a BMI of 35 kg/m$^2$, 30 years of age, and with no previous births.

**Figure 2.** Line plot showing the average (mean and 95% CI) daily activity in minutes by gestation week and treatment group for the 27 patients who completed the study. Note missing activity data in the app-coach group between 30 and 34 weeks.




## Step Counts

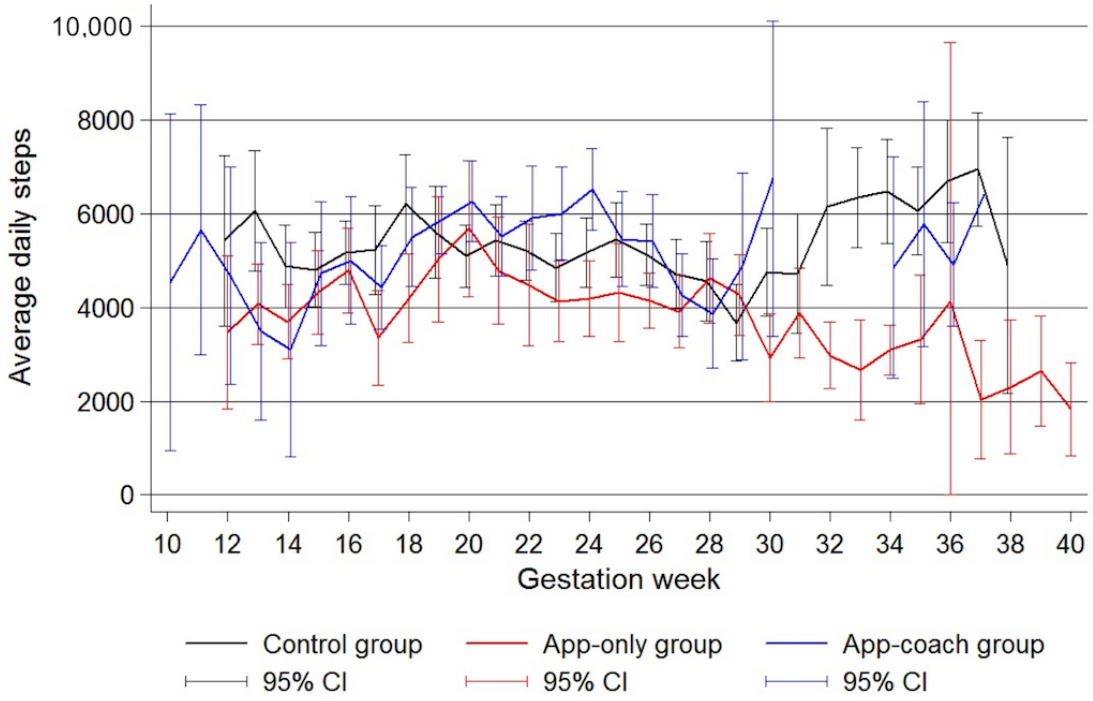
The estimated mean baseline daily step count for a 30-year-old nulliparous control patient, with a BMI of 35 kg/m$^2$ and between 61 and 90 gestational days was 5455 steps (Table 3). Gestation day was the only variable with a statistically significant effect on step count (decrease of 7.80 steps/day for each additional day of pregnancy; 95% CI 2.91 to 12.69, *P*=.002), with no difference in daily step counts between groups. From the 12th to the 29th gestational week, daily step counts did not vary between groups. However, a divergence in daily step trajectories was subsequently observed, with the average daily step count decreasing for the app group compared with participants in either the control or app-coach group, although these differences were nonsignificant (Figure 3). Overall, mean daily step counts in all groups remained in the sedentary (5000 steps/day) or low activity (5000-7499 steps/day) categories for the entire study duration. A step count of over 10,000 daily steps was recorded on 62 days over the study duration, 15 days by 4 control patients, 9 days by 3 app group patients, and 38 days by 6 app-coach patients.

**Table 3.** Results of the statistical model investigating steps.

| Variable | Coefficient | Standard error | *P* value | 95% CI |
|---|---|---|---|---|
| App group | −267.02 | 610.68 | .66 | −1463.94 to 929.90 |
| App-coach group | 661.45 | 784.76 | .40 | −876.65 to 2199.55 |
| Gestational day | −7.80 | 2.50 | .002 | −12.69 to −2.91 |
| BMI at recruitment | −90.40 | 71.65 | .21 | −230.84 to 50.03 |
| Age at recruitment | 131.94 | 82.29 | .11 | −29.35 to 293.23 |
| Parity | −276.92 | 367.83 | .45 | −997.85 to 444.01 |
| Baseline[a] | 5454.89 | 450.22 | reference | 4572.48 to 6337.30 |

[a]Baseline represents baseline step count for an individual in the control group between 61 and 90 gestational days with a BMI of 35 kg/m$^2$, 30 years of age, and with no previous births.

**Figure 3.** Line plot showing the average (mean) daily steps by gestation week and treatment group for the 27 patients who completed the study with the corresponding 95% CI. Note missing step data in the app-coach group between 30 and 34 weeks.




### Gestational Weight Gain

Mean (SD) GWG was 13.22 (5.91), 7.91 (4.17), and 13.21 (5.73) kg in control, app group, and app-coach group patients, respectively. When allowing for the increased weight at baseline of app group patients, there was no significant difference between groups in GWG, although accounting for the small sample size and resultant significant uncertainty around this estimate, the direction of effect was toward a reduction in weight gained. The results of the multiple linear regression model are shown in Table 4. An increase in weight at recruitment of 1 kg was associated with a further increase in GWG of 0.89 kg (95% CI 0.72 to 1.06 kg, *P*<.001).

**Table 4.** Results of the statistical model investigating gestational weight gain.

| Variable | Coefficient | Standard error | *P* value | 95% CI |
|---|---|---|---|---|
| App group | −5.46 | 2.84 | .07 | −11.51 to 0.59 |
| App-coach group | −0.40 | 2.80 | .89 | −6.37 to 5.57 |
| Weight at recruitment | 0.89 | 0.08 | <.001 | 0.72 to 1.06 |
| Age at recruitment | 0.28 | 0.26 | .30 | −0.27 to 0.84 |
| Parity | −3.50 | 1.64 | .050 | −7.00 to 0.01 |
| Baseline[a] | 111.17 | 2.49 | reference | 105.87 to 116.47 |

[a]Baseline represents weight at delivery for a participant in the control group with a weight at recruitment of 95 kg, age at recruitment of 30 years, and no previous births.

## Discussion

### Principal Findings

This randomized, controlled feasibility trial has demonstrated the feasibility of activity data syncing with a personal smartphone in a cohort of pregnant women with obesity. Challenges were demonstrated, however, in the delivery of the behavioral intervention and feasibility of aggregating data because of patient noncompliance with pedometer wearing and loss of devices. These resulted in high missing data rates. Inactivity was common, with baseline activity rates less than half the recommended 30 min/day. Higher BMI in early pregnancy was associated with lower activity levels, as seen in past studies, and increased GWG.

### Relationship to Prior Literature

We observed a comparable level of inactivity with other antenatal populations using pedometer data. A 2011 Australian study examined activity levels of 30 overweight or pregnant women with obesity between 26 and 28 weeks of gestation, reporting a mean (SD) daily step count of 4680 (2520) steps/day [15]. The same group, in the 2014 *HeLP-her* trial, observed a baseline mean (SD) step count of 5438 (3145) steps/day in 98 women at 12 to 15 weeks of gestation [14]. A 2010 Danish study of 338 pregnant women measured comparatively higher overall mean step counts, although lower in women with obesity:

6482, 7446, and 4626 steps/day versus 7558, 8865, and 6289 steps/day in normal-weight women at gestational week 13, 21, and 37, respectively [16]. We observed mean daily step counts significantly lower than these levels, more in keeping with prior Australian studies. Such a difference is possibly related to the high educational levels reported in the Danish study population with potentially increased activity rates. We also observed stable step counts across the gestational period. The reasons for this are unclear but are likely related to already-sedentary baseline activity levels in early pregnancy in our cohort.

Our study differed from previous pedometer-based activity interventions in the novel methodology of automated data-upload, with much-improved data integrity and completeness (overall mean days with complete data ranging between groups from 60.5% to 78.9% of all study days) compared with previous study methods. A randomized follow-up intervention by the aforementioned Danish group, for example, randomly allocated 425 pregnant women with obesity to increased pedometer-guided physical activity, compared with standard antenatal care [27]. In this study, step counts were self-recorded and reported; thus, only half of the participants reported any step count data and only for one-quarter of the study period. The Australian HeLP-her trial was similarly constrained, with pedometers being periodically worn for a specified period only (3-7 days) and the generated data extrapolated to estimate total physical activity [14]. Although our study provides a much more comprehensive picture of activity throughout pregnancy, we also observed considerable missing data rates because of patients forgetting to wear the pedometer and loss of the device, both likely related to the small pedometer size. There is thus the opportunity for future protocol refinement, such as incorporating modern smartphones directly, which have inbuilt ability to measure activity data. Recent studies have demonstrated the validity of these devices in measuring step counts, which may lead to even greater data completeness given the likely improved compliance with carrying and reduced chance of losing a personal smartphone [28,29].

We observed no difference in step counts, daily active minutes, or weight gain reduction between groups, although this feasibility trial was underpowered to assess this. Evidence suggests a major challenge is improving outcomes in a pregnant cohort already with obesity. A 2015 Cochrane review of dietary and exercise interventions in 11,000 women across 49 randomized controlled trials consistently found benefits for women with normal BMI, but no significant reduction in pregnancy weight gain for overweight or women with obesity [7]. A future definitive trial will have to overcome these challenges, likely through achieving greater engagement with a behavioral change program than was seen in this initial trial. More flexibility in the delivery of the intervention, timing to coincide with regular scheduled antenatal appointments, or batch-delivery in a group setting are all strategies that could be explored in improving engagement in a definitive trial. Our cohort also had more obesity than in the aforementioned HeLP-her trial, with a mean baseline BMI of 35.9, 36.7, and 37.0 kg/m$^2$ among groups, versus 30.3 and 30.4 in control and intervention groups in the HeLP-her trial, respectively. This may explain the disparate finding in our study, of an association with baseline BMI and increased GWG, compared with an inverse correlation in the HeLP-her trial. The significantly reduced activity levels found with increasing BMI in our cohort may explain this difference, illustrating further challenges if comparable magnitudes of obesity are observed in a definitive follow-up study population.

## Implications of the Study Findings

The findings of this study imply that wearing a pedometer with the ability to sync data with a personal smartphone is feasible in pregnant women with obesity. These findings also imply that future studies seeking to improve physical activity via wearable devices should focus on ways to improve compliance and engagement with the intervention. This study found no benefit to an additional individualized behavioral intervention from clinicians, although lower than expected engagement makes this conclusion uncertain, and this study was not powered to assess this.

## Strengths and Limitations

Strengths of this study include the novel study design using the combination of cheap and robust wearable devices with mobile phones, which are ubiquitous in a younger participant cohort in contemporary Australian society (the overall population smartphone ownership in 2018 was 89%) [30]. An additional strength of this study was the ability to successfully deliver the intervention to those women who engaged with the process. A major limitation of our study was missing data, potentially making conclusions around activity, step data, and GWG endpoints less precise. Encouragingly, data syncing and upload from patients' smartphones did not appear to be a factor. Rather, the combination of lack of pedometer wearing and outright loss of the pedometer were major contributors, all likely related to the pedometer's small size and lack of integration in patients' daily habits. Refining our trial methodology to step-counts measured directly by the newer generation of smartphones, with inbuilt activity apps, would be beneficial. Although phones may be similarly affected by noncarrying time, it is likely that there would be less overall missing data using these devices. An alternative strategy would be the use of *reminders* that are native to the smartphone operating systems or embedded within the fitness apps to prompt participants to sync data more regularly. Another strategy would be to use such reminders to improve compliance with pedometer wearing, which could be extended to direct researcher-participant contact in the event of identified poor compliance.

A further limitation was our stratification of missing data by censoring at 1000 steps/day, although there is no accepted, validated definition in the mobile health literature for what constitutes a day without appropriate pedometer usage. We note that <1000 steps/day has been used by previous studies to define a nonvalid pedometer day [31]. Other studies have used activity time as a surrogate marker of wearing, defining days with <3 hours of data recorded as missing, and 3 to 8 hours of data as half-days [14]. This approach, however, has the potential to erroneously exclude sedentary periods (during which time activity data are not being recorded, despite the pedometer being worn effectively). We could have enhanced our missing data

analysis with pedometer wearing diaries, although we note that compliance with these has been shown to be poor [14]. Complete case analysis was also used, meaning observations that contained missing data were not used. This could potentially have biased results, though we have no reason to believe that missing data rates were not distributed randomly among groups or across the gestational period.

There are also potential inaccuracies inherent in wearable pedometers, although we note a recent systematic review of Fitbit pedometers commented favorably on the measurement of steps in adults with no mobility restrictions, as in our cohort [32]. The same review did caution against inaccuracies in the active minutes measured, with a tendency to underestimating sedentary time. We also did not collect information on diet and calorie intake, which may have influenced overall GWG, although we have no reason to believe that this varied between groups. A final major limitation was the difficulty in delivering the behavioral change intervention, with only 4 of 10 women following through to intervention completion. This resulted in significantly reduced group separation, and limits the generalizability of our findings, with the possibility that true differences because of either intervention were not revealed in this feasibility trial.

## Conclusions

This study suggests that activity data syncing with a personal smartphone is feasible in a cohort of pregnant women with obesity, although our results do not support a follow-up study with this design. A future definitive study seeking to reduce GWG and improve activity in this population must focus on improving compliance with activity data recording and with the behavioral intervention delivered. Greater flexibility in intervention delivery for patients and improvements in activity monitoring through direct use of participant smartphones are strategies to explore before a definitive trial.

## Conflicts of Interest

None declared.

## References


1. McIntyre HD, Gibbons KS, Flenady VJ, Callaway LK. Overweight and obesity in Australian mothers: epidemic or endemic? Med J Aust 2012 Feb 20;196(3):184-188. [doi: 10.5694/mja11.11120] [Medline: 22339524]
2. The Royal Australian and New Zealand College of Obstetricians and Gynaecologists. 2016. Management of Obesity in Pregnancy URL: https://ranzcog.edu.au/RANZCOG_SITE/media/RANZCOG-MEDIA/Women%27s%20Health/Statement%20and%20guidelines/Clinical-Obstetrics/Management-of-obesity-(C-Obs-49)-Review-March-2017.pdf?ext=.pdf [accessed 2019-01-26]
3. Goldstein RF, Abell SK, Ranasinha S, Misso M, Boyle JA, Black MH, et al. Association of gestational weight gain with maternal and infant outcomes: a systematic review and meta-analysis. J Am Med Assoc 2017 Jun 6;317(21):2207-2225 [FREE Full text] [doi: 10.1001/jama.2017.3635] [Medline: 28586887]
4. Institute of Medicine, National Research Council Committee to Reexamine IOM. Pregnancy weight guidelines. In: Rasmussen KM, Yaktine AL, editors. Weight Gain During Pregnancy: Reexamining The Guidelines. Washington, DC: National Academy Press; 2009.
5. Cedergren MI. Optimal gestational weight gain for body mass index categories. Obstet Gynecol 2007 Oct;110(4):759-764. [doi: 10.1097/01.AOG.0000279450.85198.b2] [Medline: 17906006]
6. Callaway LK, Prins JB, Chang AM, McIntyre HD. The prevalence and impact of overweight and obesity in an Australian obstetric population. Med J Aust 2006 Jan 16;184(2):56-59. [Medline: 16411868]
7. Muktabhant B, Lawrie TA, Lumbiganon P, Laopaiboon M. Diet or exercise, or both, for preventing excessive weight gain in pregnancy. Cochrane Database Syst Rev 2015 Jun 15(6):CD007145. [doi: 10.1002/14651858.CD007145.pub3] [Medline: 26068707]
8. Ehrlich SF, Sternfeld B, Krefman AE, Hedderson MM, Brown SD, Mevi A, et al. Moderate and vigorous intensity exercise during pregnancy and gestational weight gain in women with gestational diabetes. Matern Child Health J 2016 Jun;20(6):1247-1257 [FREE Full text] [doi: 10.1007/s10995-016-1926-z] [Medline: 26955997]
9. Jiang H, Qian X, Li M, Lynn H, Fan Y, Jiang H, et al. Can physical activity reduce excessive gestational weight gain? Findings from a Chinese urban pregnant women cohort study. Int J Behav Nutr Phys Act 2012 Feb 9;9:12 [FREE Full text] [doi: 10.1186/1479-5868-9-12] [Medline: 22321640]
10. World Health Organisation. 2011. Global Recommendations on Physical Activity for Health: 18–64 Years Old URL: https://www.who.int/dietphysicalactivity/physical-activity-recommendations-18-64years.pdf?ua=1 [accessed 2020-01-26]
11. Harrison CL, Brown WJ, Hayman M, Moran LJ, Redman LM. The role of physical activity in preconception, pregnancy and postpartum health. Semin Reprod Med 2016 Mar;34(2):e28-e37 [FREE Full text] [doi: 10.1055/s-0036-1583530] [Medline: 27169984]
12. Physical Activity Australia. 2019. Pregnancy and Exercise URL: https://www.physicalactivityaustralia.org.au/pregnancy-and-exercise/ [accessed 2020-03-17]
13. Tudor-Locke C, Bassett DR. How many steps/day are enough? Preliminary pedometer indices for public health. Sports Med 2004;34(1):1-8. [doi: 10.2165/00007256-200434010-00001] [Medline: 14715035]

14. Harrison CL, Lombard CB, Teede HJ. Limiting postpartum weight retention through early antenatal intervention: the HeLP-her randomised controlled trial. Int J Behav Nutr Phys Act 2014 Oct 31;11:134 [FREE Full text] [doi: 10.1186/s12966-014-0134-8] [Medline: 25358909]
15. Harrison CL, Thompson RG, Teede HJ, Lombard CB. Measuring physical activity during pregnancy. Int J Behav Nutr Phys Act 2011 Mar 21;8:19 [FREE Full text] [doi: 10.1186/1479-5868-8-19] [Medline: 21418609]
16. Renault K, Nørgaard K, Andreasen KR, Secher NJ, Nilas L. Physical activity during pregnancy in obese and normal-weight women as assessed by pedometer. Acta Obstet Gynecol Scand 2010 Jul;89(7):956-961. [doi: 10.3109/00016341003792459] [Medline: 20583938]
17. Dodd JM, Grivell RM, Crowther CA, Robinson JS. Antenatal interventions for overweight or obese pregnant women: a systematic review of randomised trials. Br J Obstet Gynaecol 2010 Oct;117(11):1316-1326 [FREE Full text] [doi: 10.1111/j.1471-0528.2010.02540.x] [Medline: 20353459]
18. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 3;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
19. Darvall JN, Parker A, Story DA. Feasibility and acceptability of remotely monitored pedometer-guided physical activity. Anaesth Intensive Care 2016 Jul;44(4):501-506 [FREE Full text] [doi: 10.1177/0310057X1604400415] [Medline: 27456182]
20. FitBit. 2019. What Are Active Minutes? URL: https://help.fitbit.com/articles/en_US/Help_article/1379 [accessed 2020-03-17]
21. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
22. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
23. Australian Government Department of Health. 2019. Australia's Physical Activity and Sedentary Behaviour Guidelines and the Australian 24-Hour Movement Guidelines URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines [accessed 2020-03-17]
24. Eat For Health. 2019. Australian Dietary Guidelines 1-5 URL: https://www.eatforhealth.gov.au/guidelines/australian-dietary-guidelines-1-5 [accessed 2019-01-26]
25. The Royal Women's Hospital. 2019. Active Pregnancy URL: https://www.thewomens.org.au/health-information/pregnancy-and-birth/a-healthy-pregnancy/active-pregnancy [accessed 2020-11-22]
26. Renault K, Nørgaard K, Secher NJ, Andreasen KR, Baldur-Felskov B, Nilas L. Physical activity during pregnancy in normal-weight and obese women: compliance using pedometer assessment. J Obstet Gynaecol 2012 Jul;32(5):430-433. [doi: 10.3109/01443615.2012.668580] [Medline: 22663312]
27. Renault KM, Nørgaard K, Nilas L, Carlsen EM, Cortes D, Pryds O, et al. The treatment of obese pregnant women (TOP) study: a randomized controlled trial of the effect of physical activity intervention assessed by pedometer with or without dietary intervention in obese pregnant women. Am J Obstet Gynecol 2014 Feb;210(2):134.e1-134.e9. [doi: 10.1016/j.ajog.2013.09.029] [Medline: 24060449]
28. Amagasa S, Kamada M, Sasai H, Fukushima N, Kikuchi H, Lee I, et al. How well iPhones measure steps in free-living conditions: cross-sectional validation study. JMIR Mhealth Uhealth 2019 Jan 9;7(1):e10418 [FREE Full text] [doi: 10.2196/10418] [Medline: 30626569]
29. Höchsmann C, Knaier R, Eymann J, Hintermann J, Infanger D, Schmidt-Trucksäss A. Validity of activity trackers, smartphones, and phone applications to measure steps in various walking conditions. Scand J Med Sci Sports 2018 Jul;28(7):1818-1827. [doi: 10.1111/sms.13074] [Medline: 29460319]
30. Deloitte. 2019. Mobile Consumer Survey 2019 URL: https://www2.deloitte.com/au/mobile-consumer-survey [accessed 2020-03-05]
31. Houser NE, Donald SW, Kolen AM. The impact of valid pedometer days on average daily steps and wear time in children. J Phys Act Res 2019 Jan 17;4(2):103-107. [doi: 10.12691/jpar-4-2-4]
32. 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]


## Abbreviations

**GWG:** gestational weight gain

*Edited by G Eysenbach; submitted 20.06.19; peer-reviewed by K Renault, S Amagasa, J Most; comments to author 31.10.19; revised version received 23.11.19; accepted 12.02.20; published 26.05.20.*

*Please cite as:*
*Darvall JN, Wang A, Nazeem MN, Harrison CL, Clarke L, Mendoza C, Parker A, Harrap B, Teale G, Story D, Hessian E*
*A Pedometer-Guided Physical Activity Intervention for Obese Pregnant Women (the Fit MUM Study): Randomized Feasibility Study*
*JMIR Mhealth Uhealth 2020;8(5):e15112*
*URL: http://mhealth.jmir.org/2020/5/e15112/*
*doi:10.2196/15112*
*PMID:32348280*

Original Paper

# Assessment of the Efficacy of a Mobile Phone–Delivered Just-in-Time Planning Intervention to Reduce Alcohol Use in Adolescents: Randomized Controlled Crossover Trial

Severin Haug[1], PhD; Raquel Paz Castro[1], PhD; Urte Scholz[2], PhD; Tobias Kowatsch[3,4], PhD; Michael Patrick Schaub[1], PhD; Theda Radtke[5], PhD

[1]Swiss Research Institute for Public Health and Addiction, University of Zurich, Zurich, Switzerland

[2]Applied Social and Health Psychology, University of Zurich, Zurich, Switzerland

[3]Center for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland

[4]Center for Digital Health Interventions, Department of Technology, Management, and Economics, ETH Zurich, Zurich, Switzerland

[5]Health, Work & Organizational Psychology, School of Psychology and Psychotherapy, Witten/Herdecke University, Witten/Herdecke, Germany

**Corresponding Author:**
Severin Haug, PhD
Swiss Research Institute for Public Health and Addiction
University of Zurich
Konradstrasse 32
Zurich, 8005
Switzerland
Phone: 41 444481174
Email: severin.haug@isgf.uzh.ch

## Abstract

**Background:** Interventions to reduce alcohol use typically include several elements, such as information on the risks of alcohol consumption, planning for sensible drinking, and training of protective behavioral strategies. However, the effectiveness of these individual intervention elements within comprehensive programs has not been addressed so far, but it could provide valuable insights for the development of future interventions. Just-in-time interventions provided via mobile devices are intended to help people make healthy decisions in the moment and thus could influence health behavior.

**Objective:** The aim of this study was to test the proximal effects of a mobile phone–delivered, just-in-time planning intervention to reduce alcohol use in adolescents who reported recent binge drinking. The efficacy of this individual intervention element was tested within a comprehensive intervention program to reduce problem drinking in adolescents.

**Methods:** The study had an AB/BA crossover design, in which participants were randomly allocated to (1) a group receiving the planning intervention (A) in period 1 and assessment only (B) in period 2 or (2) a group receiving assessment only (B) in period 1 and the planning intervention (A) in period 2. The planning intervention included a text message to choose one of two predetermined if-then plans to practice sensible drinking with friends or when going out and a prompt to visualize the chosen plan. There was a washout period of at least 1 week between period 1 and period 2.

**Results:** Out of 633 program participants who recently binge drank, 136 (21.5%) were receptive in both periods of time and provided data on the proximal outcome, which was the number of alcoholic drinks consumed with friends or when going out. After the planning intervention, the number of alcoholic drinks consumed was approximately one standard drink lower compared with the finding without the intervention (P=.01).

**Conclusions:** A mobile phone–delivered, just-in-time, if-then planning intervention to practice sensible drinking with friends or when going out is effective in reducing alcohol consumption among adolescents who report recent binge drinking. Based on the relatively low percentage of participants with self-reported receptivity for the planning intervention, measures to increase the population impact of similar planning interventions should be implemented and tested in future trials.

**Trial Registration:** ISRCTN Registry ISRCTN52150713; http://www.isrctn.com/ISRCTN52150713

***(JMIR Mhealth Uhealth 2020;8(5):e16937)*** doi:10.2196/16937

**KEYWORDS**

alcohol; adolescents; planning intervention; just-in-time intervention; crossover trial

## Introduction

Alcohol use is a major cause of disease burden in most countries worldwide and is among the 10 leading risk factors in all Central European countries [1]. In young people, drinking is associated with multiple social and interpersonal problems, such as arguing with friends and parents, engaging in unplanned sexual activity, drinking and driving, assault, getting into trouble with the law, academic difficulties, unintended injuries, and suicidal acts [2,3]. In the long term, individuals with problematic alcohol use exhibit an elevated risk of developing chronic conditions, such as heart and liver diseases and alcohol use disorders.

Internationally recognized indicators of problem drinking are (1) average daily consumption of more than two standard drinks for men and one standard drink for women [4] and (2) binge drinking, which is defined as drinking at least five standard drinks on a single occasion for men and four drinks on a single occasion for women [5]. In particular, binge drinking prevalence rates are high in adolescence and young adulthood. In Switzerland, the binge drinking prevalence on a monthly basis is 25% in adolescents aged 15 to 19 years and 41% in young adults aged 20 to 24 years [6]. The prevalence of elevated mean daily consumption in young people is low (2% at 15–19 years of age and 8% at 20–24 years of age) relative to binge drinking, and it almost always occurs in combination with binge drinking [6].

Interventions, including personalized normative feedback and drinking reduction strategies, as major intervention elements show small short-term effects on the reduction of binge drinking prevalence in young people [2]. These intervention elements are typically included in comprehensive intervention programs, which include several elements derived from major psychological models of health behavior change, such as social norms, outcome expectations, motivation, self-efficacy, and planning interventions. Planning interventions, including *if-then plans*, are among the most recognized and frequently applied planning techniques adopted to change health behavior [7]. These strategies, also known as implementation intentions, require people to specify a critical situation and pair it with a goal-directed behavioral response (if situation *x* occurs, then I will show behavior *y*). Behavioral responses could be self-generated or prespecified, as in the intervention of this study. Laboratory research showed that both prespecified and self-generated if-then planning interventions were effective to reduce alcohol use in a representative sample of adults [8] and in alcohol-consuming adolescents [9]. Beyond laboratory research, the proximal effects of specific intervention elements (so called microinterventions like implementation intentions) might also be tested within comprehensive intervention programs. Testing the effects of these microinterventions within traditionally delivered comprehensive intervention programs allows balancing internal and external validities in a way that facilitates translation and testing of the basic theory in multicomponent intervention programs [10]. According to this approach, the effect of the intervention element could be studied within real-life conditions and assignment of participants to a control group without any intervention is not necessary. Unlike traditional methods of delivering planning interventions, mobile phones can deliver these microinterventions “just in time” for when a person is most vulnerable and receptive. These just-in-time interventions can be activated by users themselves (user triggered) through prespecified rules (server triggered, like in this study) or sensors that dynamically monitor a user’s context (context triggered) [11]. They are intended to support an individual at the time when most needed [12]. To date, published studies on the effects of just-in-time interventions are limited to physical activity and sedentary behavior, with mixed evidence for intervention effects [13].

In this study, we tested the proximal effects of a mobile phone–delivered, just-in-time, if-then planning intervention to reduce alcohol use in adolescents who reported recent binge drinking. We hypothesized that the number of alcoholic drinks consumed with friends or when going out would be lower during the just-in-time planning intervention as compared with assessment only.

## Methods

### Study Objectives and Design

The study aimed to determine the proximal effects of a just-in-time planning intervention to reduce alcohol use in adolescents who reported recent binge drinking. The study was registered at Current Controlled Trials International Standard Randomized Controlled Trials Number (ISRCTN 52150713, assigned June 2, 2017). The study protocol was approved by the ethics committee of the Philosophical Faculty at the University of Zurich, Switzerland (date of approval: April 18, 2017). The trial was executed in compliance with the Declaration of Helsinki.

Within a randomized controlled AB/BA crossover design, each participant received the planning intervention (A) and assessment only (B) in a randomized order. The trial was conducted in Switzerland, and participants were recruited between June 2017 and July 2018. Participants were recruited in vocational and upper secondary schools and participated in a comprehensive mobile phone–based intervention program to reduce problem drinking with a duration of 3 months. The inclusion criteria were (1) ownership of a mobile phone, (2) recent binge drinking, (3) alcohol consumption in the evening/night with friends or when going out, and (4) available data on preferred if-then plans. The just-in-time planning intervention was based on effective implementation intention and action planning interventions [7]. On two of their typically indicated drinking days, participants either received (A) a text message to choose one of two predetermined if-then plans to practice sensible drinking with friends or when going out and subsequently another text message prompting to visualize the chosen plan or (B) no intervention. There was a washout period of at least 1 week between A and B as well as B and A.

### Procedure and Participants of the Comprehensive Intervention Program

Vocational and upper secondary schools in the Swiss cantons of Zurich and Berne were invited to participate in the comprehensive mobile phone–based intervention program named *MobileCoach Alcohol* by prevention specialist centers. Sixteen vocational and upper secondary schools, with 108 classes in total, agreed to participate in the program and the study. Research assistants (psychology master's degree students or graduates) invited all of the students in the participating classes to take part in an online health survey during a regular school lesson reserved for health education. Online screening was conducted using tablet computers provided by the research assistants or the students' own smartphones. Demographic data, alcohol consumption, smoking status, and mobile phone ownership were assessed. The only inclusion criterion for participation in the comprehensive program was ownership of a mobile phone. A total of 1710 students were present in the school classes. Out of these, 1676 participated in the online screening and 1419 (83.0% of the students present in the classes) consented to participate in the *MobileCoach Alcohol* intervention program and provided their mobile phone number. Program participants were informed about data protection, the aims of the program and study, assessments, and reimbursement. Informed consent was obtained online from all program and study participants. The automated intervention program included online feedback provided immediately after the baseline assessment and individually tailored text messages provided over 3 months. The online feedback included normative feedback based on the social norms approach [14]. The text messages for binge drinking participants focused on (1) motivation to drink within low-risk limits, using individual data concerning positive outcome expectancies [15]; (2) alcohol-related problems, established using individual data on previous alcohol-related problems; (3) peak blood alcohol concentration and related risk calculated using data concerning sex, body weight, and maximum number of drinks consumed on a single occasion in the preceding month; and (4) strategies to resist alcohol when going out or when being with friends. Additionally, three text message assessments were performed during the intervention period. First, a quiz on the metabolism of alcohol, for which participants received immediate individualized feedback on their answers, and if they did not respond within 48 hours, they were sent the correct responses. Second, a contest that required participants to create a text message to motivate other participants to drink within low-risk limits. The best text message, rated weekly by an alcohol prevention specialist from the Swiss Research Institute for Public Health and Addiction, was sent anonymously to all other participants after 48 hours. Third, an assessment of binge drinking within the preceding week, which included immediate individualized feedback. The text messages typically contained 150 to 300 characters and were delivered via SMS text messaging. Several text messages also included web links to thematically appropriate video clips, pictures, and websites. Except for the text messages providing the planning intervention, which were typically sent on Fridays and Saturdays at 5 pm (mentioned below), messages were typically sent on Tuesdays at 6 pm.

### Procedure and Participants of the Planning Intervention Study

The participants for this study to test the proximal effects of the mobile phone–delivered, just-in-time, alcohol planning intervention were selected automatically by a computer algorithm within the *MobileCoach Alcohol* intervention program (Figure 1).

The computer algorithm selected recent binge drinking adolescents with alcohol consumption in the evening or night when going out or when being with friends and with available data on preferred if-then plans. A total of 633 (44.6%) of the 1419 *MobileCoach Alcohol* intervention program participants fulfilled the inclusion criteria and were randomly allocated to (1) a group receiving the planning intervention (A) in period 1 and assessment only (B) in period 2 or (2) a group receiving assessment only (B) in period 1 and the planning intervention (A) in period 2. The randomization sequence with a 1:1 ratio was created using computerized random numbers.

The assessment questions and the text messages of the planning intervention are depicted in Figure 2.

The data necessary for the provision of the just-in-time planning intervention were assessed within the baseline assessment. This assessment was also performed during the school lesson immediately after the online screening and informed consent procedure for the comprehensive intervention program (mentioned above). It included (1) selection of the typical drinking day in the course of the week when going out or when being with friends, (2) selection of the typical drinking time when going out or when being with friends, and (3) selection of two out of nine favorite if-then plans providing strategies to drink little or no alcohol with friends or when going out. One reason to select only two out of nine if-then plans was that these plans were also presented within SMS text messages, which are restricted in their length and number of characters. Another reason was not to overstrain the participant within this situation. Regarding the typical drinking days, we did not include Monday, Tuesday, and Wednesday, as previous studies [16,17] on this program showed that practically nobody chose these days. Furthermore, we wanted to prevent temporal overlap of the just-in-time intervention elements (presented Thursday to Sunday) and the other intervention elements of the program, which were typically sent on Tuesdays. The text messaging–based part included (1) assessment of the state of receptivity on the individually indicated drinking day at 5 pm, (2) the actual planning intervention comprising a text message to choose one of two predetermined if-then plans to practice sensible drinking with friends or when going out and another text message prompting to visualize the chosen plan, and (3) assessment of the proximal outcome (the number of alcoholic drinks consumed on the previous day with friends or when going out). The assessment-only condition solely included (1) and (3). The planning intervention was designed considering the latest recommendations for research and practice on planning and implementation intentions in health contexts [7].

As the entire intervention program had a total duration of 12 weeks and we considered a washout period of 1 week as appropriate between period 1 and 2, assessments of the state of

receptivity and potential provision of the subsequent planning intervention were possible six times during the intervention (in weeks 1, 3, 5, 7, 9, and 11). In order to obtain a maximum sample size for this crossover trial, the state of receptivity was assessed in as many weeks as possible until a participant was receptive twice, that is, there were up to six chances for the period 1 assessment, and the remaining assessments after responding to the period 1 assessment were for period 2. After being receptive twice, the participants no longer received this state of the receptivity assessment. The time interval for responding to the state of the receptivity assessment was 6 hours. Participants who did not respond within this time period and those who indicated that they did not meet with friends or go out did not receive the subsequent messages of the planning intervention and the outcome assessment.

**Figure 1.** Flow of study participants.

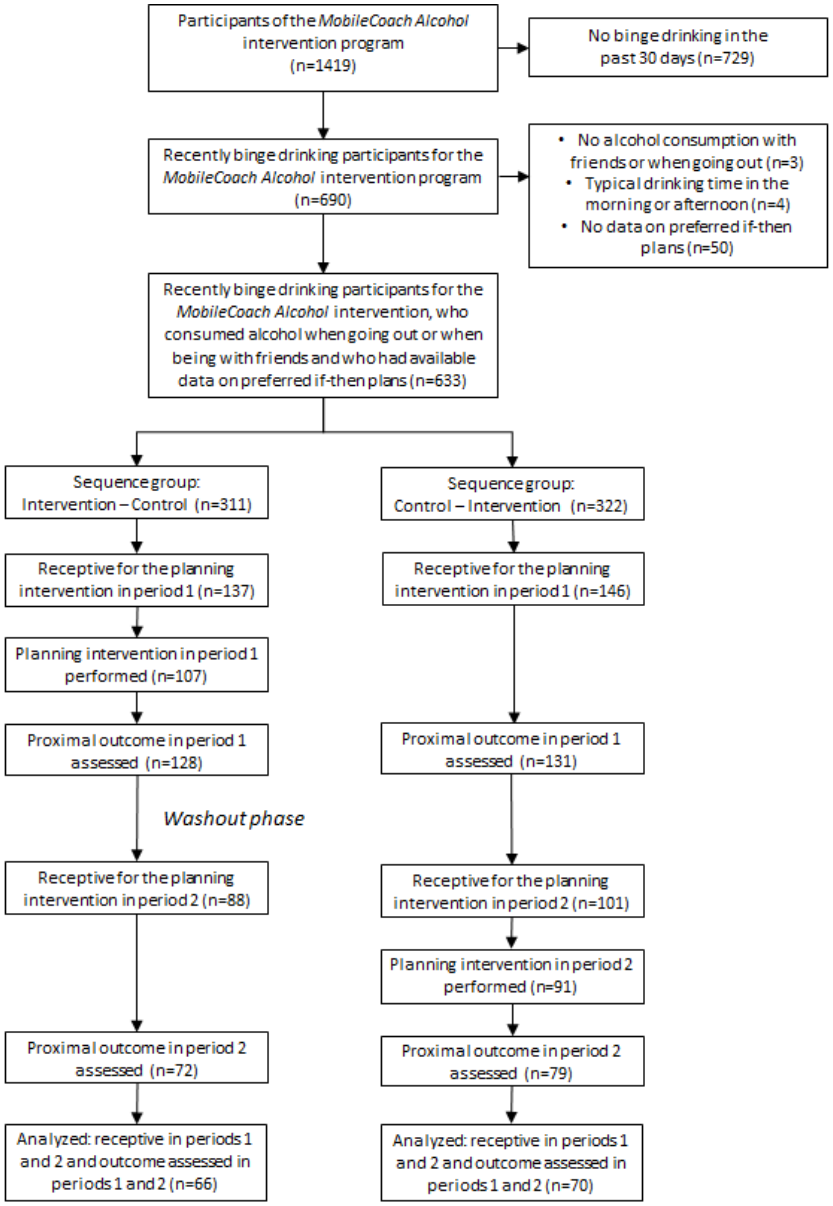
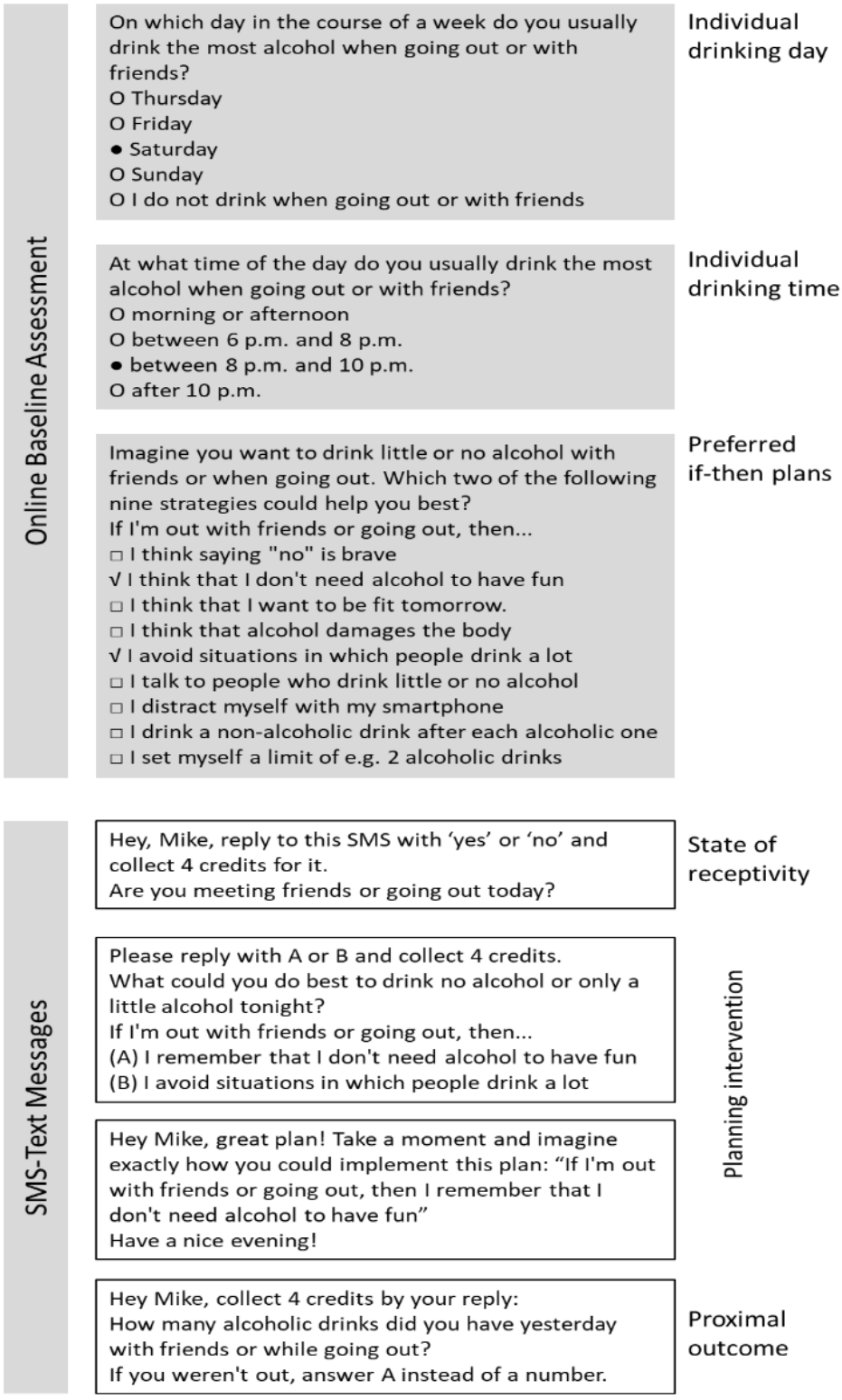


**Figure 2.** Assessments and planning intervention from participants' perspective.




## Sample Size Calculation

The estimation of the effect size was based on the results of a controlled study testing the effectiveness of implementation intentions to reduce alcohol use in a sample of the general population [8], using an online power calculator for crossover studies provided by the Biostatistics Center of Massachusetts General Hospital [18]. Within this study [8], the participants randomized to the experimenter-provided implementation intention condition were presented with a choice of three implementation intentions, from which they chose the one they thought would work best for them and wrote it down. Participants with high alcohol intake under this condition reduced their alcohol intake from baseline to a 1-month follow-up by 1.3 standard drinks per day, whereas this reduction was 0.1 drinks in the passive control group. Using this minimal detectable difference in means of 1.2, a standard deviation within

participants of 3.0, a power of 80%, and a noncentral *t* function (α=5%, one-sided), the study required a total of 79 participants.

### Assessments and Outcomes

The online baseline assessment included the following variables: sex, age, immigration background, tobacco smoking, the Alcohol Use Disorders Identification Test (AUDIT) [19], and binge drinking. We assessed the countries of birth for students' parents to identify a potential immigrant background. Based on this information, participants were assigned to one of the following categories: (1) neither parent born outside Switzerland, (2) one parent born outside Switzerland, and (3) both parents born outside Switzerland. In the analysis, we combined one- and two-sided immigrant backgrounds into a single category and compared it with a nonimmigrant background.

Tobacco smoking was assessed using the following question: "Do you currently smoke cigarettes or have you smoked in the past?" The response options were as follows: (1) I smoke cigarettes daily; (2) I smoke cigarettes occasionally but not daily; (3) I smoked cigarettes in the past, but I do not smoke anymore; and (4) I have never smoked cigarettes or have smoked less than 100 cigarettes throughout my life. In the analysis, we combined categories (1) and (2) as smokers and categories (3) and (4) as nonsmokers. Alcohol use was assessed through the consumption items of the AUDIT (AUDIT-C) [19]. The AUDIT-C assesses drinking quantity, drinking frequency, and binge drinking frequency, and it has a potential range from 0 to 12, with higher values representing higher alcohol use. Binge drinking prevalence in the preceding 30 days was assessed by asking participants to report the number of standard drinks consumed on the heaviest drinking occasion in the preceding 30 days. Examples of standard drinks containing 12 to 14 g of ethanol were provided for beer, wine, spirits, alcopops, and cocktails, along with conversion values (eg, three 0.5 L cans of beer equals six standard drinks). Binge drinking was defined as drinking at least five drinks on a single occasion in males and four drinks on a single occasion in females [5].

The primary outcome of this planning intervention study was the number of alcoholic drinks with friends or when going out on the day of the intervention. This proximal outcome was assessed 24 hours after assessment of the state of receptivity, that is, at 5 pm on the day following the individual indicated drinking day. The secondary outcome of this planning intervention study was binge drinking, which was defined as drinking at least five drinks in males and four drinks in females with friends or when going out on the individual indicated drinking day.

### Data Analysis

We initially examined the data of the primary outcome, which was based on the self-reported number of alcoholic drinks entered as free text. Based on a visual inspection of the distributions and the recommendations of Osborne and Overbay [20], outliers were identified at more than 3 standard deviations above the mean and adjusted to 3 standard deviations above the mean.

We used chi-square tests for categorical variables and *t* tests for continuous variables to evaluate baseline differences between the analyzed sample and the sample of participants not responding or receptive in both periods of time in order to examine the representativeness of the sample.

Intervention effects of the planning intervention were tested following the recommendations of Wellek and Blettner [21] on proper data analysis in clinical trials with a crossover design. These recommendations include the following: (1) participants are assigned randomly to the two sequence groups AB and BA; (2) the crucial variable for analysis is the within-subject difference in outcome between the two study periods, tested by a valid test for independent samples comparing the values of this variable between the sequence groups; and (3) an assumption that the washout phase is long enough to rule out a carryover effect, which should be checked by another test for independent samples.

Following the latter recommendation and to rule out that the treatment effects were confounded by time effects or a carryover effect, as the washout phase was not long enough, we calculated the sum of the number of drinks consumed in the two periods for each subject and compared it across the two sequence groups by a *t* test for independent samples.

To test for the effects of the alcohol planning intervention, we used a *t* test for independent samples for the primary outcome (number of alcoholic drinks on the previous day with friends or when going out) and a chi-square test for the binary secondary outcome (binge drinking on the previous day with friends or when going out). All outcome analyses were based on a complete-case dataset, which included participants who were receptive in periods 1 and 2 and had outcomes assessed in both time periods. Results with a type I error rate involving a *P* value <.05 in two-sided tests were considered statistically significant, with the exception of the proximal outcome that was tested one-sided, as the study hypotheses and power calculations were also based on one-sided tests. Analyses were performed using SPSS, version 22 (IBM Corp, Armonk, New York, USA).

## Results

### Study Participation

Figure 1 depicts the participants' progression through the trial. A total of 633 binge drinking participants for the *MobileCoach Alcohol* intervention, who consumed alcohol in the evening or night when going out or when being with friends and who had available data on preferred if-then plans, were randomly allocated to the sequence group AB (n=311) or the sequence group BA (n=322). Out of these 633 participants, 136 (21.5%) were receptive twice and provided data on the proximal outcome at both time points. Our analyses on the effectiveness of the just-in-time alcohol planning intervention were based on these 136 participants.

This analyzed group did not differ from participants who could not be assessed (n=497) with respect to the following baseline characteristics: age, immigration background, tobacco smoking status, AUDIT-C, typical drinking day, and typical drinking time on the indicated drinking day. However, the nonreceptive participants or nonresponders were more likely male (51.5% vs

40.8%, $\chi^2_1$=4.9; *P*=.03) and in upper secondary than vocational schools (45.5% vs 32.4%; $\chi^2_1$=7.5; *P*=.006).

Characteristics of the analyzed study sample (n=136) and the two sequence groups (AB/BA) are shown in Table 1.

**Table 1.** Baseline characteristics of the study sample.

| Variable | Sequence: intervention–control (AB) (n=66) | Sequence: control–intervention (BA) (n=70) | Total (n=136) |
|---|---|---|---|
| **Sex, n (%)** | | | |
| Male | 35 (53.0%) | 35 (50.0%) | 70 (51.5%) |
| Female | 31 (47.0%) | 35 (50.0%) | 66 (48.5%) |
| Age, mean (SD) | 16.9 (1.0) | 17.2 (1.3) | 17.1 (1.1) |
| **Immigration background, n (%)** | | | |
| No immigration background | 30 (45.5%) | 42 (60.0%) | 72 (52.9%) |
| One or both parents born outside Switzerland | 36 (54.5%) | 28 (40.0%) | 64 (47.1%) |
| **Type of school, n (%)** | | | |
| Upper secondary school | 20 (30.3%) | 24 (34.3%) | 44 (32.4%) |
| Vocational school | 46 (69.7%) | 46 (65.7%) | 92 (67.6%) |
| **Tobacco smoking status, n (%)** | | | |
| Daily or occasional cigarette smoking | 31 (47.0%) | 34 (48.6%) | 65 (47.8%) |
| Nonsmoking | 35 (53.0%) | 36 (51.4%) | 71 (52.2%) |
| AUDIT-C[a], mean (SD) | 6.2 (1.6) | 6.7 (1.7) | 6.4 (1.7) |
| **Typical drinking day, n (%)** | | | |
| Thursday | 1 (1.5%) | 0 (0%) | 1 (0.7%) |
| Friday | 28 (42.4%) | 27 (38.6%) | 55 (40.4%) |
| Saturday | 37 (56.1%) | 43 (61.4%) | 80 (58.8%) |
| Sunday | 0 (0%) | 0 (0%) | 0 (0%) |
| **Typical drinking time on the indicated typical drinking day, n (%)** | | | |
| Morning or afternoon | 0 (0%) | 0 (0%) | 0 (0%) |
| Between 6 pm and 8 pm | 4 (6.1%) | 5 (7.1%) | 9 (6.6%) |
| Between 8 pm and 10 pm | 32 (48.5%) | 40 (57.1%) | 72 (52.9%) |
| After 10 pm | 30 (45.5%) | 25 (35.7%) | 55 (40.4%) |

[a]AUDIT-C: consumption items of the Alcohol Use Disorders Identification Test

## Test for Period Effects

To rule out that the treatment effects were confounded by time effects or carryover effects, we calculated the sum of the number of drinks consumed in the two periods for each subject and compared it across the two sequence groups by a *t* test for independent samples. This test did not suggest that any time or period effects were present (*t*=−1.19, *P*=.24).

## Effects of the Alcohol Planning Intervention

As shown in Table 2, in the AB and BA sequence groups, the mean numbers of alcoholic drinks consumed on the previous day with friends or when going out were 2.79 and 3.43, respectively, after receiving the planning intervention and 3.68 and 4.21, respectively, without the intervention. The within-subject differences of 0.89 and −0.79 between the sequence groups were statistically significant (*t*=2.31, *P*=.01), showing that the planning intervention is effective to reduce the mean alcohol use on typical drinking days in young people by about one standard drink.

Concerning the secondary outcome (binge drinking on the previous day with friends or when going out), in the AB and BA sequence groups, the prevalences were 30% (20/66) and 36% (25/70), respectively, after receiving the planning intervention and 35% (23/66) and 36% (25/70), respectively, without the intervention. A chi-square test comparing the number of participants who did not change between period 1 and period 2, who reported binge drinking in period 2 but not in period 1, and who reported binge drinking in period 1 but not in period 2, showed no statistically significant difference between the two sequence groups ($\chi^2_2$=1.34, *P*=.25).

**Table 2.** Effects of the alcohol planning intervention.

| Variable | Sequence: intervention–control (AB) (n=66) | | | Sequence: control–intervention (BA) (n=70) | | | Test value | *P* value |
|---|---|---|---|---|---|---|---|---|
| | Period 1 (A) | Period 2 (B) | Within-subject difference (period 2−period 1) | Period 1 (B) | Period 2 (A) | Within-subject difference (period 2−period 1) | | |
| Number of alcoholic drinks on the previous day with friends or when going out, mean (SD) | 2.79 (3.09) | 3.68 (3.62) | 0.89 (3.71) | 4.21 (3.67) | 3.43 (3.85) | −0.79 (4.69) | 2.31[a] | .01 |
| Binge drinking on the previous day with friends or when going out, n (%) | 20 (30%) | 23 (35%) | 3 (5%) | 25 (36%) | 25 (36%) | 0 (0.0%) | 1.34[b] | .25 |

[a]*t* test for independent samples for the comparison of the within-subject difference between the condition sequences.

[b]Chi-square test for the comparison of binge drinking change in period 2 compared with period 1 between the condition sequences.

## Discussion

This study aimed to test the proximal effects of a just-in-time planning intervention for reducing alcohol use in adolescents who reported recent binge drinking. To the best of our knowledge, this is the first study to test the effects of a just-in-time intervention for reducing alcohol consumption. The study revealed the following three main results, which are discussed below: (1) It is feasible to test the proximal effects of single intervention elements like implementation intentions within a comprehensive multicomponent intervention program if the program is delivered via a mobile phone and has a minimum duration; (2) An if-then alcohol planning intervention is effective to reduce the mean alcohol use on typical drinking days in young people by about one standard drink; (3) A large proportion of adolescents is not receptive to the just-in-time planning intervention, when the state of receptivity is assessed via text messaging and a reply within a limited response time is required for triggering the intervention.

This study underlines that the proximal effects of specific intervention elements like just-in-time implementation intentions could be tested in a randomized controlled crossover design within a comprehensive intervention program, if some requirements are met as follows: (1) The intervention program is provided automatically (eg, via a mobile phone that allows server-triggered just-in-time interventions); (2) The duration of the program is long enough to assess receptivity for the intervention several times in order to have at least two time points for comparison; and (3) The intervention element is presented after a long enough time (wash-out period) from the other elements or contents of the program in order to reduce confounding. Another requirement, which was hardly met in previous trials on just-in-time interventions, is an adequate sample size to have enough statistical power [13]. Currently published trials in the areas of physical activity and sedentary behavior typically have much smaller sample sizes than this trial, although the state of receptivity and proximal outcomes were often assessed automatically and unobtrusively via smartphone sensors or activity trackers [22].

The result that a single if-then planning intervention is effective to reduce alcohol consumption is in line with previous findings derived from laboratory research [8,9] and extends the insights to that effect that the planning intervention could also be provided digitally without any personal contact, as well as just-in-time on the individually indicated drinking day. Owing to different follow-up intervals, the achieved reduction in alcohol use could not directly be compared with other interventions. However, to enable a rough estimation, a traditional if-then planning intervention in adolescents resulted in an average reduction of 2.5 g of alcohol per day (one-fifth of a standard drink) at a 2-month follow-up [9]. An earlier version of the digital *MobileCoach Alcohol* intervention program resulted in an average reduction of 0.4 standard drinks per day at a 6-month follow-up [17].

Although the just-in-time delivery of the alcohol planning intervention might be partly responsible for its effectiveness, it remains unclear whether the population impact (number of participants reached multiplied by effectiveness [23]) of the provided planning intervention might have been greater if selection of the preferred if-then plan would have taken place within a longer timeframe before the drinking day (eg, the whole day before) or an initially preferred if-then plan would have been sent irrespective of the state of receptivity. Both options may be able to increase the impact of this planning intervention as compared with the procedure implemented in this trial, where an active reply on receptivity via text messaging within a limited response time was required. However, direct comparisons of different delivery options or qualitative participant interviews are necessary for a better understanding of the limited reach. Additional analyses of the replies to the receptivity question “Are you meeting friends or going out today?” showed that a substantial part of the lack of receptivity was due to the denial of this question; in weeks 1 and 3 when all study participants received this question, 231 out of 492 (47.0%) denied meeting friends or going out that day. These data suggest that half of adolescents are not receptive because they are not at risk for binge drinking on a predetermined day. Correspondingly, measures to update or adapt the drinking day during program delivery might be promising to increase receptivity.

Several limitations of this study should be mentioned. First, the number of alcoholic drinks consumed was self-reported, and in contrast to the baseline assessment, no examples of standard drinks were provided within the text messaging–based assessment of the proximal outcome. Second, although the crossover design applied has several advantages (eg, it avoids problems of comparability of the study and control groups because each participant is his/her own control and the required sample size is low), carryover effects might have confounded part of the intervention effects, although the respective finding was not relevant. Third, the sample analyzed within this study systematically differed from all participants randomized, with respect to sex and type of school, which limits the external validity of the results. Fourth, we solely used baseline data for tailoring the intervention time. The possibility of updating or adapting the intervention time periodically might increase intervention effectiveness.

In conclusion, this study shows that just-in-time interventions could be tested and implemented in the area of addiction and that digitally provided alcohol planning interventions could reduce alcohol use in adolescents who report recent binge drinking. Future studies should focus on increasing the reach and outcome of just-in-time alcohol planning interventions by testing other delivery formats or by sensor-triggered intervention delivery, which, for example, dynamically monitors a participant's context and provides support when high-risk environments, such as areas with many nightlife locations and social situations, are sensed [24,25].

## Acknowledgments

Funding for this project was provided by the Swiss Federal Office of Public Health (grant no. 16.929761).

## Conflicts of Interest

TK is affiliated with the Center for Digital Health Interventions (www.c4dhi.org), a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. TK is also the cofounder of Pathmate Technologies, a university spinoff company that creates and delivers digital clinical pathways and has used the open source MobileCoach platform for that purpose. The funding institution did not influence the design and conduct of the study; the management, analysis or interpretation of data; or the preparation, review or approval of the manuscript.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 391 KB - mhealth_v8i5e16937_app1.pdf ]

## References

1. GBD 2015 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1659-1724 [FREE Full text] [doi: 10.1016/S0140-6736(16)31679-8] [Medline: 27733284]
2. Kuntsche E, Kuntsche S, Thrul J, Gmel G. Binge drinking: Health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):976-1017. [doi: 10.1080/08870446.2017.1325889] [Medline: 28513195]
3. Kuntsche E, Gmel G. Alcohol consumption in late adolescence and early adulthood--where is the problem? Swiss Med Wkly 2013;143:w13826 [FREE Full text] [doi: 10.4414/smw.2013.13826] [Medline: 23888405]
4. U.S. Department of Health and Human Services and U.S. Department of Agriculture. U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015. 2015–2020 Dietary Guidelines for Americans, 8th Edition URL: https://health.gov/sites/default/files/2019-09/2015-2020_Dietary_Guidelines.pdf [accessed 2020-04-27]
5. Gmel G, Kuntsche E, Rehm J. Risky single-occasion drinking: bingeing is not bingeing. Addiction 2011 Jun;106(6):1037-1045. [doi: 10.1111/j.1360-0443.2010.03167.x] [Medline: 21564366]
6. Gmel G, Kuendig H, Notari L, Gmel C. Suchtmonitoring Schweiz: Konsum von Alkohol, Tabak und illegalen Drogen in der Schweiz im Jahr 2015 [Addiction monitoring Switzerland: use of alcohol, tobacco and illegal drugs in Switzerland, data 2015]. Lausanne: Sucht Schweiz; 2016. URL: https://www.suchtmonitoring.ch/docs/library/gmel_thezckxvvkq4.pdf [accessed 2020-04-27]
7. Hagger MS, Luszczynska A, de Wit J, Benyamini Y, Burkert S, Chamberland P, et al. Implementation intention and planning interventions in Health Psychology: Recommendations from the Synergy Expert Group for research and practice. Psychol Health 2016 Jul;31(7):814-839. [doi: 10.1080/08870446.2016.1146719] [Medline: 26892502]
8. Armitage CJ. Effectiveness of experimenter-provided and self-generated implementation intentions to reduce alcohol consumption in a sample of the general population: a randomized exploratory trial. Health Psychol 2009 Sep;28(5):545-553. [doi: 10.1037/a0015984] [Medline: 19751080]
9. Armitage CJ, Rowe R, Arden MA, Harris PR. A brief psychological intervention that reduces adolescent alcohol consumption. J Consult Clin Psychol 2014 Jun;82(3):546-550. [doi: 10.1037/a0035802] [Medline: 24491079]

10. Strauman TJ, Socolar Y, Kwapil L, Cornwell JF, Franks B, Sehnert S, et al. Microinterventions targeting regulatory focus and regulatory fit selectively reduce dysphoric and anxious mood. Behav Res Ther 2015 Sep;72:18-29 [FREE Full text] [doi: 10.1016/j.brat.2015.06.003] [Medline: 26163353]
11. Naughton F. Delivering "Just-In-Time" Smoking Cessation Support Via Mobile Phones: Current Knowledge and Future Directions. Nicotine Tob Res 2017 Mar;19(3):379-383. [doi: 10.1093/ntr/ntw143] [Medline: 27235703]
12. Liao P, Klasnja P, Tewari A, Murphy SA. Sample size calculations for micro-randomized trials in mHealth. Stat Med 2016 May 30;35(12):1944-1971 [FREE Full text] [doi: 10.1002/sim.6847] [Medline: 26707831]
13. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
14. Perkins H. The Social Norms Approach to Preventing School and College Age Substance Abuse: A Handbook for Educators, Counselors, and Clinicians. San Francisco: Wiley, John & Sons, Incorporated; 2003.
15. Babor TF, Higgins-Biddle JC. Brief Intervention for Hazardous and Harmful Drinking: A Manual for use in Primary Care. Geneva: World Health Organization: General Department of Mental Health and Substance Dependence; 2001.
16. Haug S, Schaub MP, Venzin V, Meyer C, John U, Gmel G. A pre-post study on the appropriateness and effectiveness of a Web- and text messaging-based intervention to reduce problem drinking in emerging adults. J Med Internet Res 2013 Sep 02;15(9):e196 [FREE Full text] [doi: 10.2196/jmir.2755] [Medline: 23999406]
17. Haug S, Paz Castro R, Kowatsch T, Filler A, Dey M, Schaub MP. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in adolescents: Results of a cluster-randomized controlled trial. J Consult Clin Psychol 2017 Feb;85(2):147-159. [doi: 10.1037/ccp0000138] [Medline: 27606700]
18. Massachusetts General Hospital Biostatistics Center. Statistical considerations for a cross-over study where the outcome is a measurement URL: http://hedwig.mgh.harvard.edu/sample_size/js/js_crossover_quant.html [accessed 2020-04-25]
19. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [doi: 10.1001/archinte.158.16.1789] [Medline: 9738608]
20. Osborne JW, Overbay A. The power of outliers (and why researchers should always check for them). Practical Assessment, Research, and Evaluation 2004;9(6) [FREE Full text] [doi: 10.7275/qf69-7k43]
21. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2012 Apr;109(15):276-281 [FREE Full text] [doi: 10.3238/arztebl.2012.0276] [Medline: 22567063]
22. Klasnja P, Smith S, Seewald N, Lee A, Hall K, Luers B, et al. Efficacy of Contextually Tailored Suggestions for Physical Activity: A Micro-randomized Optimization Trial of HeartSteps. Ann Behav Med 2019 May 03;53(6):573-582 [FREE Full text] [doi: 10.1093/abm/kay067] [Medline: 30192907]
23. Thyrian JR, John U. Population impact--definition, calculation and its use in prevention science in the example of tobacco smoking reduction. Health Policy 2007 Aug;82(3):348-356. [doi: 10.1016/j.healthpol.2006.10.001] [Medline: 17126947]
24. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
25. Kuntsche E, Labhart F. The future is now--using personal cellphones to gather data on substance use and related factors. Addiction 2014 Jul;109(7):1052-1053. [doi: 10.1111/add.12540] [Medline: 24754712]


## Abbreviations

**AUDIT:** Alcohol Use Disorders Identification Test
**AUDIT-C:** consumption items of the Alcohol Use Disorders Identification Test


*Edited by G Eysenbach; submitted 06.11.19; peer-reviewed by MY Chih, J Thrul, C Meyer; comments to author 14.12.19; revised version received 27.01.20; accepted 24.03.20; published 26.05.20.*

*Please cite as:*
*Haug S, Paz Castro R, Scholz U, Kowatsch T, Schaub MP, Radtke T*
*Assessment of the Efficacy of a Mobile Phone–Delivered Just-in-Time Planning Intervention to Reduce Alcohol Use in Adolescents: Randomized Controlled Crossover Trial*
*JMIR Mhealth Uhealth 2020;8(5):e16937*
*URL: http://mhealth.jmir.org/2020/5/e16937/*
*doi:10.2196/16937*
*PMID:32452818*

Original Paper

# Effect of an mHealth Intervention Using a Pedometer App With Full In-Person Counseling on Body Composition of Overweight Adults: Randomized Controlled Weight Loss Trial

Alberto Hernández-Reyes[1*], MSc, RD; Fernando Cámara-Martos[1*], PhD; Rafael Molina-Luque[1*], MSc; Rafael Moreno-Rojas[1*], PhD

Universidad de Córdoba, Córdoba, Spain

[*]all authors contributed equally

**Corresponding Author:**
Alberto Hernández-Reyes, MSc, RD
Universidad de Córdoba
Darwin Building, Annex
Campus de Rabanales
Córdoba, 14014
Spain
Phone: 34 687376563
Email: z52heloa@uco.es

## Abstract

**Background:** In clinical practice, it is difficult to convey the benefits of sustained physical activity to adult patients with excess weight or obesity. For this purpose, a goal-setting walking prescription may be an effective strategy.

**Objective:** This study aimed to determine the efficacy of the intervention of a pedometer app in setting a goal to reach 10,000 steps per day in adults.

**Methods:** Overweight adults (n=98; mean body mass index 32.53 [SD 4.92] kg/m2) were randomized to one of two conditions (control or intervention). Both groups downloaded a pedometer app that recorded their daily step counts and were given a daily walking goal of 10,000 steps. Subjects participated in a 24-week in-person behavioral weight control program and were asked to monitor their daily levels using the pedometer app. Baseline data were recorded and followed up weekly. Only the intervention group had structured information delivery, a personalized physical activity prescription, and follow-up on number of steps per day.

**Results:** The results show that regardless of sex or age, prescribing walking increased the number of steps per day by 4806 step on average (standardized β coefficient=–0.813, SE=427.586, $t$=–11.242, $P$<.001).

**Conclusions:** These results could have implications for improving self-monitoring in overweight adults during periods of weight loss. Health professionals should analyze the implementation of tools that permit them to prescribe, follow up, and encourage the achievement of a goal of physical activity in overweight or obese patients.

**Trial Registration:** ClinicalTrials.gov NCT03845478; https://clinicaltrials.gov/ct2/show/NCT03845478

***(JMIR Mhealth Uhealth 2020;8(5):e16999)*** doi:10.2196/16999

**KEYWORDS**

pedometer; physical activity; exercise prescription; diet; health behavior

## Introduction

Three in four adolescents and one in three adults in the world do not perform at least 30 minutes of physical activity (PA) per day as recommended by the World Health Organization [1]. This represents a severe public health problem, given that physical inactivity is the fourth leading risk factor for global mortality and plays a role in the development of obesity and metabolic syndrome [2,3]. Increasing PA should be a priority in the treatment of weight loss in overweight and obese subjects. Walking is a solution to overcoming physical inactivity [4] due to its low impact, in which the person can control its intensity, duration, and effort in order to reduce the risk of injury [5]. A recent review has deepened the knowledge of the psychosocial

factors that increase adherence to a PA protocol, and social support and self-evaluation are presented as critical elements [6]. The prescription of PA by the health professional to sedentary people suffering from obesity, diabetes, or hypertension is defined as the delivery of an individualized exercise prescription for a limited time [6].

Despite efforts in many countries to facilitate access to leisure centers so people can increase PA [7], lack of adherence remains the main problem [8].

Monitoring PA by counting steps per day offers the possibility of standardizing evaluation and follow-up [9]. Although the goal of 10,000 steps a day may not be appropriate for all ages and levels of physical conditioning, it is considered a reasonable and motivating goal for healthy adults, and previous studies have demonstrated its effectiveness in weight loss programs [9,10].

The term mHealth is defined as "public health and medical practices compatible with mobile devices, including smart mobile phones, patient monitoring gadgets, personal digital assistants (PDA), and other wireless devices" [11]. The use of this technology is pervasive, and, in developed countries, the level of penetration reaches almost 100% of the adult population [11]. Mobile apps provide tools, processes, and communications used to support and provide medical care to patients and the general public [12]. Apps monitoring PA are objective and automated, allowing the user to carry a device that tracks their movements [13].

Different methods are available to measure PA, but there is no definite gold standard to measure PA in various clinical settings. Accelerometers and pedometers have been developed for research use because they are easy to wear and portable. Accelerometers can provide quantification and recording of PA [14]. A important issue with accelerometers is how to select cutoff points to define activity intensities. Despite proposed cutoffs for some devices, there is currently no consensus [15], and this inconsistency in the use of accelerometers to delineate exercise makes it difficult to compare findings of different studies [16].

Pedometers are capable of counting the number of steps; they became popular more than a decade ago as a meter gauge and motivator of daily exercise [17]. Their use today has grown significantly thanks to the development of apps capable of collecting and storing information on daily PA concerning walking or running. This technology has been proven to be effective in the strategy to encourage and motivate patients to execute and even increase the number of steps per day [18]. The review by Mansi et al [19] concluded that interventions based on the use of a pedometer were more effective when combined with additional behavioral strategies (eg, setting goals and facilitating access to the information generated).

The main objective of this study was to evaluate the efficacy of prescribing versus recommending PA in a sedentary adult population with overweight or obesity. Another aim of the study was to measure the improvement of body composition in both scenarios. This study sought to extend the findings of Glynn et al [20] by examining the feasibility of this approach of using the Accupedo-Pro pedometer app (Corusen LLC) intervention to promote PA in an overweight adult sample. The hypothesis is that encouraging, following, and involving patients in the use of this app through the establishment of objectives and self-monitoring would significantly increase the number of steps executed per day, thus affecting the amount and quality of weight loss, and we can measure this by comparing total body weight, body fat and overall muscle mass, and body mass index (BMI).

## Methods

### Recruitment

Participants were recruited from private health clinics and sports centers through social network advertisements and direct actions in the centers in the area of Cádiz, Andalusia, Spain. The exclusion and inclusion criteria are listed in Textbox 1. Participants were interested in losing weight and owned a smartphone. Participants attended an orientation session to complete a consent form and baseline questionnaires, including the International Physical Activity Questionnaire and demographic questions. The study protocol complied with the Declaration of Helsinki for medical studies, it was approved by the bioethical committee of Córdoba University and the department of health at the regional government of Andalusia (Act no. 284, ref. 4156), and the study was retrospectively registered with ClinicalTrials.gov [NCT03845478] on February 19, 2019.

**Textbox 1.** Selection criteria.

**Inclusion:**

- Aged 18 years and older
- Overweight or obese (body mass index 25 to 49.9 kg/m$^2$)
- Own smartphone with Android or iOS operating system and internet access
- Lead sedentary lifestyle
- Have not been on a diet to lose weight within 6 months of the start of the study

**Exclusion:**

- Diabetes treated with oral medications or insulin
- Pregnant
- Chronic renal insufficiency

## Randomization Groups

The number of daily reference steps counted was collected in all the subjects using the Accupedo app for 7 days before randomization. Patients did not receive any information or comments at the time of app installation. Subsequently, subjects were randomly assigned (1:1) using a computerized random number generator, and Accupedo was installed on phones in both groups. In their initial interview, members of the two groups received information on the importance and benefits of walking 10,000 steps a day, but only the intervention group (IG) received follow-up and monitoring to reach this total. Patients in both groups were instructed to use the pedometer daily during all their waking hours. Each week, a member of the research team checked the data on participant apps and recorded the average daily steps taken during the week and month in the computer system. The IG had individualized goal settings, instructions on counting steps for self-assessment, and educational and motivational content to improve self-management. The control group (CG) received a recommendation to count steps, without any reproach in the case of not increasing their PA.

## Push Notifications

Push notifications permit the delivery of timely updates and customized reminders to users. This functionality offers auditory and visual alerts to inform users about an incoming message and invites them to act, even if the app sending the notifications is not currently in use. The methodology we used is explained in greater detail elsewhere [21].

## Outcome Assessments

### *Physical Activity*

After randomization, the number of daily steps was averaged each week if the subject's pedometer was used on three or more days during that week. The difference between the average daily step count and the baseline one was determined for weeks 12 to 24, inclusive. The strata proposed by Tudor-Locke and Bassett were used [9], establishing the following ranges based on the evidence available for classifying PA according to the data provided by the pedometer:

- Fewer than 5000 steps per day can be used as a sedentary lifestyle index
- 5000 to 7499 steps per day is considered not very active
- 7500 to 9999 probably includes some volitional activities (or high demands for occupational activity) and could be considered active
- 10,000 steps or more per day indicates the point used to classify people as quite active

### *Dietary Intervention*

The daily energy requirements were determined by estimating the resting energy expenditure using the formula proposed by Harris-Benedict [22]:

- Women: basal metabolic rate = 655.1 + (9.563 × weight in kg) + (1.850 × height in cm) – (4.676 × age in years)
- Men: basal metabolic rate = 66.5 + (13.75 × weight in kg) + (5.003 × height in cm) – (6.755 × age in years)

and multiplying the value obtained by a factor of 1.5 in those patients performing PA [23].

Through 24 weeks, all patients followed a diet with the following allocation of macronutrients: 25% to 30% protein, 40% to 45% carbohydrates, and 30% to 35% fat. The menu was hypocaloric with a reduction of 500 kcal/day during the treatment period to achieve a weekly weight loss of 400 grams. After being included in the study, patients participated in a 1-hour seminar in which the dietitian-nutritionist instructed them on how to make a suitable selection of food. The menu proposed was valid for one week and was given to participants in the weekly revision appointment as the protocol for the next week. The energy and nutritional intake was evaluated by the program Dietowin and the weighing method by Dietowin 8.0 (Dietowin SL) [24].

### *Anthropometrics and Body Composition Measurements*

Body fat, muscle mass, and percentage of water, considered as result variables, were monitored by multifrequency electrical impedance (BWB-800A, Tanita Corp), which has been previously validated [25]. This method is based on a 3-compartment model capable of evaluating body fat, muscle mass, and bone mineral content. The independent variables collected were age (years), height (cm), weight (kg), and BMI

($kg/m^2$). The anthropometric measurements were taken following the recommendations of the standardized anthropometry handbook [26] by experienced personnel to reduce the coefficient of variation. Each measurement was taken 3 times, and the mean value was calculated. All quantitative variables were measured with the precision of 0.1. A stadiometer (Seca 213, Seca) was employed to measure height.

### Statistical Analysis

#### *Power Calculation*

A similar previous 6-month trial was used to conduct sample size calculations ($\alpha$=.05 and power $\beta$=80%) based on the expected differences between groups in the use of the pedometer to increase PA and body weight modification in obese adults [27].

#### *Statistics*

Quantitative variables have been presented with the mean and standard deviation, and qualitative variables in frequencies and percentages. To contrast goodness of fit with a normal distribution of data from quantitative variables, the Kolmogorov-Smirnov test with Lilliefors correction was used. For the bivariate hypothesis, the Student *t* test was performed for 2 means, while for the qualitative variables, the chi-square and Fisher exact tests were employed. For analysis of 3 or more means, analysis of variance of repeated means determined the effects of the intervention at the basal moment at 3 and 6 months, and the correlation between the quantitative variables was verified by the coefficient of Pearson correlation (*r*). Finally, if the normality or homoscedasticity criterion was not met for analysis of variance, the Kruskal-Wallis test was performed. To adjust the possible impact of PA on body composition and its possible role as a confounding factor, adjusted linear regressions were made for each body composition variable (body fat and muscle mass) and weight, calculating the standardized $\beta$ coefficients. To determine goodness of fit of the models, the standard error, adjusted coefficient of determination, *F* statistic, linearity, and residuals were analyzed. For all statistical analyses, an $\alpha$ error of less than 5% was accepted ($P<.05$) and a 95% confidence interval was calculated. Statistical analysis was performed using SPSS Statistics software version 22.0 (IBM Corp).

## Results

### Characteristics of the Population Studied

Between January 2016 and December 2018, 98 participants were randomly registered and assigned to the groups (Figure 1). No significant differences were found in the baseline data between the groups, to which the patients were assigned randomly ($P>.05$ for all; Table 1). The attrition rate was 6.12% at 3 months and 31.63% at 6 months and did not differ between groups at either 3 (chi-square = 0.33, $P=.29$) or 6 (chi-square = 0.54, $P=.09$) months.

**Figure 1.** Consolidated Standards of Reporting Trials diagram.

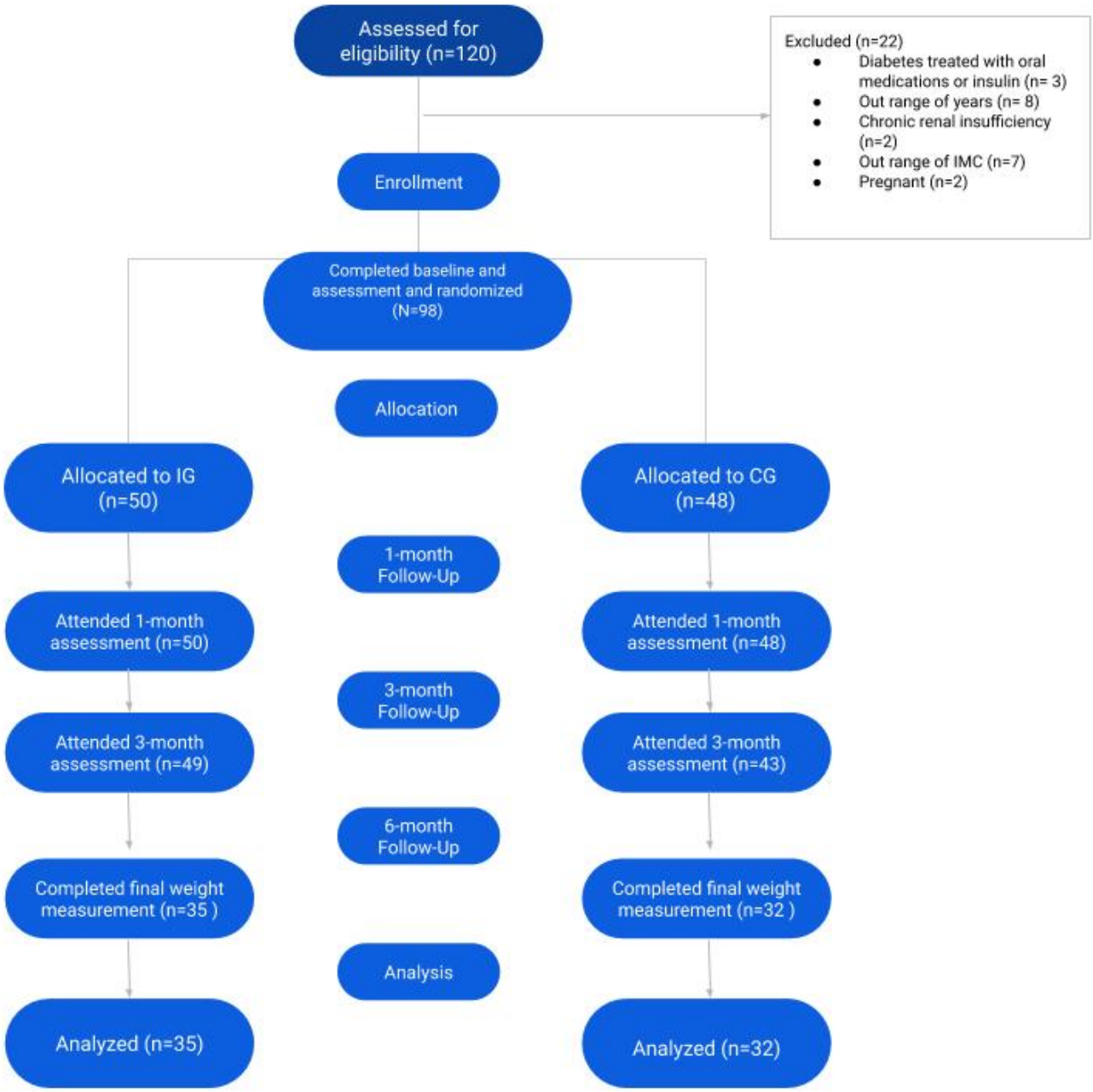


**Table 1.** Baseline body mass index and demographics of study participants.

| Variable | Total (n=67) | CG[a] (n=32) | IG[b] (n=35) | *P* value |
|---|---|---|---|---|
| Age in years, mean (SD) | 41.51 (11.29) | 42.88 (10.91) | 40.26 (11.63) | .35 |
| **Gender, n (%)** | | | | .12 |
| Female | 44 (66) | 24 (75) | 20 (57) | — |
| Male | 23 (34) | 8 (25) | 15 (43) | — |
| **Education, n (%)** | | | | .24 |
| Without studies | 17 (25) | 9 (28) | 8 (23) | — |
| Ninth grade | 25 (37) | 15 (47) | 10 (29) | — |
| High school diploma | 10 (15) | 3 (9) | 7 (20) | — |
| University students | 15 (22) | 5 (16) | 10 (29) | — |
| **Occupation, n (%)** | | | | .01 |
| Unemployed | 4 (6) | 1 (3) | 3 (9) | — |
| Service occupation | 32 (48) | 18 (56) | 14 (40) | — |
| Technical, sales, administrative | 9 (13) | 6 (19) | 3 (9) | — |
| Executive, professioanl specialty | 15 (22) | 6 (9) | 9 (34) | — |
| Retired | 4 (6) | 4 (13) | 0 (0) | — |
| Student | 3 (5) | 0 (0) | 3 (9) | — |
| Initial weight (kg), mean (SD) | — | 89.65 (19.32) | 91.35 (16.36) | .70 |
| BMI[c] ($kg/m^2$), mean (SD) | — | 32.72 (5.56) | 32.34 (4.28) | .75 |
| Body fat (%), mean (SD) | — | 39.66 (7.59) | 36.76 (8.85) | .16 |
| Muscle mass (kg), mean (SD) | — | 51.35 (13.38) | 54.92 (13.18) | .28 |
| Water (kg), mean (SD) | — | 44.10 (4.58) | 46.07 (5.66) | .13 |

[a]CG: control group.

[b]IG: intervention group.

[c]BMI: body mass index.

## Weight Modification and Body Composition Depending on Physical Activity Prescription or Not at 3 and 6 Months

Examining both groups together, participants lost significant weight at both 3 (–6.84 [SD 3.97] kgs; *P*<.001) and 6 months (–7.92 [SD 3.93] kgs; *P*<.001). Weight loss over the study period was significantly different between the groups at 3 and 6 months (Table 2). These results are extrapolated to BMI and loss of total body fat in the cited period. Change in muscle mass was not significant at 3 or 6 months. For body composition, focusing on body fat loss, prescribing PA resulted in losing fat significantly at 3 months (9.56% in IG compared with 6.13% in CG; *P*<.18) and 6 months (15.60% in IG versus 7.04% in CG; *P*<.001).

**Table 2.** Weight loss, body mass index, and body composition of participants.

| Variable | CG[a] (n=32), mean (SD) | IG[b] (n=35), mean (SD) | *P* value |
|---|---|---|---|
| **Weight change (kg)** | | | |
| 3 months | –5.63 (2.60) | –8.06 (2.59) | <.001 |
| 6 months | –6.29 (2.63) | –10.80 (3.31) | <.001 |
| **BMI[c] change (%)** | | | |
| 3 months | –5.91 (2.84) | –8.02 (2.73) | .003 |
| 6 months | –6.32 (2.79) | –10.92 (3.74) | <.001 |
| **Body fat (%)** | | | |
| 3 months | 37.38 (8.06) | 33.24 (8.32) | .02 |
| 6 months | 36.94 (7.66) | 31.12 (8.62) | <.001 |
| **Muscle mass (kg)** | | | |
| 3 months | 50.27 (12.78) | 53.54 (12.59) | .65 |
| 6 months | 50.30 (12.98) | 53.44 (12.43) | .64 |
| **Body water (%)** | | | |
| 3 months | 45.45 (5.26) | 48.41 (5.62) | .03 |
| 6 months | 46.14 (5.18) | 50.07 (5.54) | .005 |

[a]CG: control group.

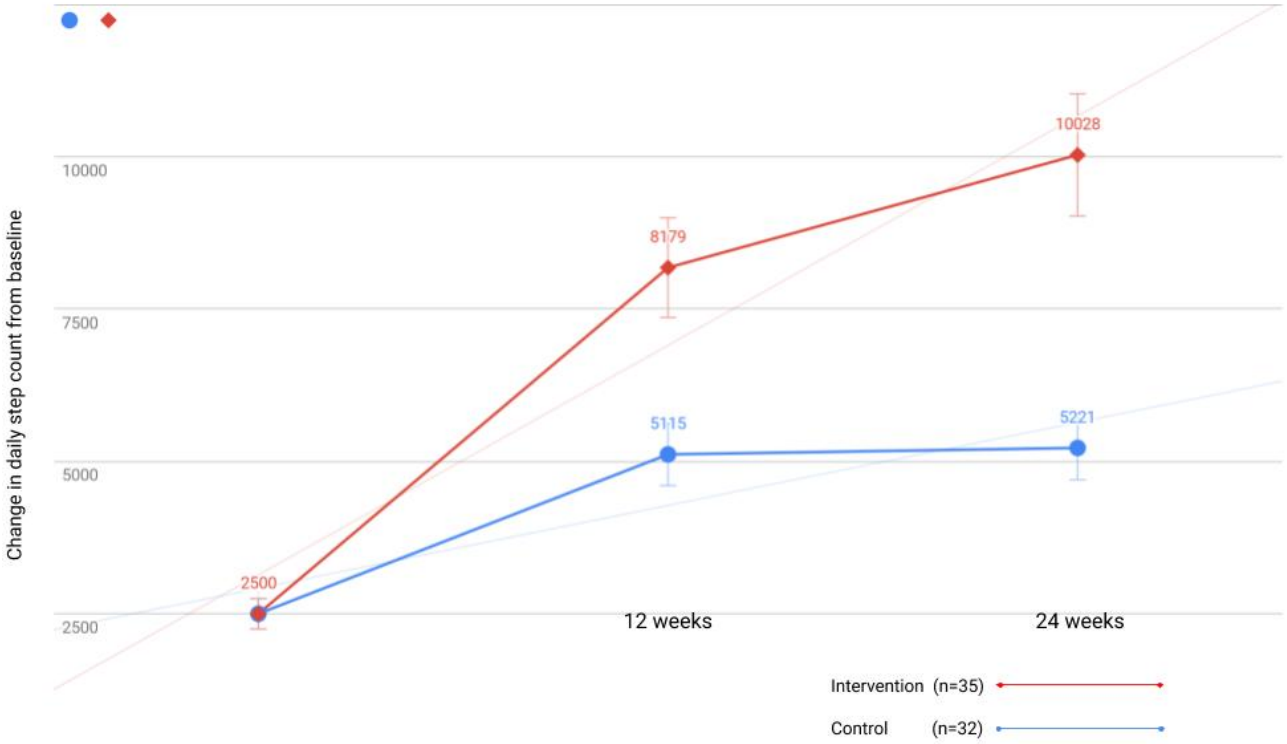
[b]IG: intervention group.

[c]BMI: body mass index.

### Increase in Average Number of Steps Depending on the Prescription or Not of Reaching 10,000 Steps a Day

Figure 2 illustrates the change for average daily steps from the start of the study to weeks 12 and 24. In week 12, subjects who only followed the recommendation to complete a certain number of steps had an average change from the baseline data of 2615 (SD 1849) steps, compared with an increase of 5679 (SD 4015) steps in the group of patients receiving the prescription. It is interesting to observe that, at 24 weeks, the CG maintained the number of daily steps, with an increase of only 106 (SD 74) steps compared with the IG that achieved 10,028 average daily steps (SD 1307). At 3 months, IG participants recorded an average of 8179.77 (SD 1815.66) steps, which is significantly higher than the average recorded by the patients in the CG (5115.25 [SD 1200.54]) steps (*P*<.001).

**Figure 2.** Graph showing interaction effect between time and condition.




## Weight Modification and Body Composition Depending on the Range of Steps Executed

Independently of the intervention or control group, we analyzed the results in weight and body composition according to the range of steps executed. The results (Table 3) show that a higher number of steps resulted in a more significant loss of body weight and BMI at 3 months ($P$<.001) and 6 months ($P$<.001), the highest BMI lost, with identical statistical results.

Focusing again on the results obtained in body fat, we found that, like weight and BMI, the difference was significant for participants in all 4 step ranges ($P$<.001). We observed that in the second period (weeks 13 to 24), patients who walked fewer than 5000 steps per day did not lose additional weight (–4.37 kgs [SD 2.3] in the first 12 weeks and –4.58 [SD 1.75] at 24 weeks). Also, in this range of steps, (<5000 per day), patients in the second period (weeks 13 to 24) began to regain the fat lost in the first 12 weeks (–6.65% [SD 4.54] in the first quarter, and –4.87% [SD 6.13] in the second). This threshold of steps per day is considered to be of concern and might be the reason for not losing weight or fat. The influence of the group assigned was endorsed by the results offered by the linear regression model, adjusted for age and sex ($R^2$ adjusted=.655, $F$=126.386; $P$<.001). This showed how, regardless of sex and age, being incorporated into the prescription group (standardized β coefficient=–0.813, SE=427.586, $t$=–11.242; $P$<.001) raised the number of steps per day at 6 months to 4806.

**Table 3.** Participant anthropometrics and body composition per step range.

| Variable | >10,000 steps, mean (SD) | 7500-9999 steps, mean (SD) | 5001-7499 steps, mean (SD) | <5000 steps, mean (SD) | *P* value |
|---|---|---|---|---|---|
| **Weight change (kg)** | | | | | |
| 3 months | –9.97 (2.48) | –8.30 (1.93) | –6.61 (2.42) | –4.37 (2.30) | <.001 |
| 6 months | –11.89 (2.88) | –10.51 (3.48) | –7.39 (2.38) | –4.58 (1.75) | <.001 |
| **BMI change (%)** | | | | | |
| 3 months | –10.45 (2.83) | –7.92 (1.58) | –6.62 (2.42) | –4.91 (3.15) | <.001 |
| 6 months | –11.97 (3.59) | –10.80 (3.66) | –7.49 (2.57) | –4.52 (1.70) | <.001 |
| **Body fat (%)** | | | | | |
| 3 months | –15.14 (7.84) | –6.92 (3.44) | –6.83 (5.41) | –6.65 (4.54) | .07 |
| 6 months | –17.13 (9.62) | –14.19 (11.45) | –9.27 (6.03) | –4.87 (6.13) | .97 |
| **Muscle mass (kg)** | | | | | |
| 3 months | –1.51 (4.86) | –3.96 (3.89) | –2.74 (2.95) | –0.08 (3.41) | .08 |
| 6 months | –2.35 (5.11) | –2.45 (4.28) | –2.18 (3.64) | –1.65 (3.78) | .02 |
| **Body water (%)** | | | | | |
| 3 months | 9.00 (6.77) | 3.29 (3.19) | 3.79 (3.31) | 2.66 (4.25) | — |
| 6 months | 9.82 (6.51) | 9.17 (6.37) | 5.52 (6.29) | 3.63 (4.03) | .05 |

[a]CG: control group.

[b]IG: intervention group.

[c]BMI: body mass index.

# Discussion

## Principal Findings

Although performing exercise regularly has been associated with the prevention of a wide range of pathologies in the developed world, the correlation between walking daily and completing a certain number of steps and its percentage quantification in weight loss, fat, and BMI is not clear. This study uses an objective measure of PA through a goal-setting mechanism and its comparison with a control group to elucidate the improvement in body composition in overweight or obese adults.

## Comparison With Prior Work

Our results show that PA prescription, together with the establishment of setting a goal for participants, improves the count of steps at 6 months compared with the use of a pedometer without a goal being fixed. We observed a decrease in the number of average daily steps during the second quarter, after a substantial initial increase. In the CG, a stagnation occurred, while in the IG, it was possible for participants to reach the goal proposed. The reminder messages sent through push notifications and work in the weekly face-to-face consultation increased the effectiveness of Accupedo. The subjects of both groups showed a better commitment, increasing their number of steps. Prescribing versus recommending, self-control, and reinforcement in face-to-face consultations seem to have been effective in promoting and maintaining the number of steps achieved daily. These results in a cohort of patients with obesity referred to a nutrition consultation, in addition to confirming previous findings in studies of similar duration and population, broaden our understanding of the beneficial mechanisms of implementing prescription strategies in a personal consultation with each patient [10,28].

The amount of weight loss in the IG can be considered clinically relevant. Helping subjects establish a realistic goal is a vital part of the success of the program. The motivation to lose weight and adding a component of PA produces a more significant fat loss and BMI adjustment in addition to increasing the degree of adherence [29].

The critical differences between the groups led us to think that, in addition to IG participants burning more calories due to the fact per se of walking more, this stimulus would cause more considerable changes in the individual. With the same diet, walking results in percentage fat losses of 15.40% in the IG compared with 7.65% in the CG. This fact may be due to a significant improvement in the regulation of the energy balance and better general functioning of the organism (ie, precise control of body homeostasis) [30]. It seems logical to propose an active lifestyle as a measure to influence the body composition of overweight or obese subjects, highlighting the impact it can have on the loss of total body fat. The results of our study confirm data obtained previously in which performing moderate PA resulted in weight and fat loss and better body composition of overweight or obese people [31,32].

We found a significant inverse relationship between the number of steps executed and BMI. Our data corroborate those obtained by Wayne et al [33] aimed at the study of this relationship in the young population, as well as those trials on middle-aged people of Thompson et al [34] or elderly ones by Krumm et al [35]. It makes sense that performing a more significant number of steps results in better body composition and BMI. However, we observed diminishing performance over time. When it comes

to making a general recommendation regarding the number of steps necessary to reduce body fat and body weight, our results confirm the range proposed by Tudor-Locke [36] (ie, maintaining a level of 8000 to 10,000 steps per day).

Our findings have implications in the field of public health. There is a positive correlation between number of steps taken and amount of body fat loss. Although decrease in BMI can be considered positive, it is a marker that today presents great controversy when interpreting its scope due to the limitations it presents [37]. However, there is consensus on the negative impact that an increase in body fat has on health as it is associated with the risk of cardiovascular disease [38], obesity [39], and an increase in general mortality [40].

Our data on steps counted show the essential objectivity of the established prescription. The review by Kang et al [41] found that pedometer use led to an increase of 2000 to 2500 steps per day, an average figure that coincides with the increment of steps achieved by the CG in our study. However, to reach 10,000 average steps per day, the pedometer per se seems to be ineffective, and some authors have suggested the need to reinforce patient behavior with self-control tools, goal setting, and follow-up to significantly increase the PA performed [42]. We corroborate these results and find them also consistent with those reported by Samdal et al [43], where a definite improvement is observed in the number of steps executed in the IG (through goal setting and in-person counseling) during the first-trimester intervention [43]. We speculate that a period longer than 12 weeks may be necessary for a sedentary and obese person to reach a volume of steps equivalent to 10,000 per day, an aspect that must be taken into account in establishing an objective when prescribing and monitoring PA.

There are several reasons why IG patients lose more weight than CG patients. First, there is self-management, a mechanism that has allowed IG participants to keep a record of self-weighing and monitor themselves to modify their behaviors to achieve goals. Another reason could be motivation and its relationship with behavior change techniques are used to encourage physical activity. The theory of motivation has been defined as being a critical mechanism in the construction process that determines intensity and direction of the action of human behavior [44,45]. The relationship with PA maintenance seemed to be influenced by the presence of an objective establishment and an active follow-up, aspects corroborated in previous trials [46].

### Limitations

Strengths of our study include randomized design, weekly control of PA in all subjects in a face-to-face consultation, demographic data, and the study of body composition. Due to the voluntary nature of the study, a limitation could be the possible self-selection occurring for highly motivated populations. We have been able to improve the degree of adherence to the prescription of PA, but we understand that 6 months, although it is a prudential period, cannot be considered definitive to confirm a change in behavior and that more time is necessary in order to confirm the effectiveness of the prescription. In addition, a 25% attrition rate was expected at 6 months, and the study saw attrition of 30%. While attrition was greater that intended, it was similar to what has been observed in other weight loss interventions [25,47]. No differential attrition rates were observed between groups in our study.

### Conclusion

Establishing goal-setting and feedback mechanisms on PA can increase the effectiveness of prescribing it in people who are overweight or obese. We are aware that the recommendations of the official bodies, the World Health Organization among others, to do PA are not fulfilled, and that people's sedentary lifestyle is a real problem. Establishing achievable objectives and developing a monitoring system capable of tracking PA and thus being able to help change sedentary habits is feasible and could be more so with the help of technology. The ease of use of the pedometer installed in the smartphone makes this tool an ally for the health care professional, who can, at no cost, have access to patient mobility data on a day-to-day basis. The application of these measures in groups of patients can be investigated in future studies. We have found advantages in short-term adherence (the first 6 months) to PA in patients with a prescription, but we do not know what happens when they leave this follow-up. Studies are needed to determine if the behavioral change persists with the passage of time or patients return to sedentary habits once they have no professional follow-up.

## Acknowledgments

We would like to thank all participants for agreeing to share their data and experiences. We would also like to thank the Laboratory Vidal-Zambrano health center for providing a comfortable work space.

## Authors' Contributions

All authors have contributed equally to the study and writing of the article. All authors have read, reviewed, and approved the final version of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2932 KB - mhealth_v8i5e16999_app1.pdf ]

## References


1. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Global Health 2018 Oct;6(10):e1077-e1086. [doi: 10.1016/s2214-109x(18)30357-7]
2. Ekkekakis P, Vazou S, Bixby WR, Georgiadis E. The mysterious case of the public health guideline that is (almost) entirely ignored: call for a research agenda on the causes of the extreme avoidance of physical activity in obesity. Obes Rev 2016 Apr;17(4):313-329. [doi: 10.1111/obr.12369] [Medline: 26806460]
3. Xiao J, Shen C, Chu MJ, Gao YX, Xu GF, Huang JP, et al. Physical activity and sedentary behavior associated with components of metabolic syndrome among people in rural China. PLoS One 2016;11(1):e0147062 [FREE Full text] [doi: 10.1371/journal.pone.0147062] [Medline: 26789723]
4. Heron N, Tully MA, McKinley MC, Cupples ME. Physical activity assessment in practice: a mixed methods study of GPPAQ use in primary care. BMC Fam Pract 2014 Jan 15;15:11 [FREE Full text] [doi: 10.1186/1471-2296-15-11] [Medline: 24422666]
5. Ferreira G, Pereira MG. Physical activity: the importance of the extended theory of planned behavior, in type 2 diabetes patients. J Health Psychol 2017 Sep;22(10):1312-1321. [doi: 10.1177/1359105315626787] [Medline: 26841795]
6. Eynon M, Foad J, Downey J, Bowmer Y, Mills H. Assessing the psychosocial factors associated with adherence to exercise referral schemes: a systematic review. Scand J Med Sci Sports 2019 May;29(5):638-650. [doi: 10.1111/sms.13403] [Medline: 30742334]
7. Arsenijevic J, Groot W. Physical activity on prescription schemes (PARS): do programme characteristics influence effectiveness? Results of a systematic review and meta-analyses. BMJ Open 2017 Feb 02;7(2):e012156 [FREE Full text] [doi: 10.1136/bmjopen-2016-012156] [Medline: 28153931]
8. Campbell F, Holmes M, Everson-Hock E, Davis S, Buckley WH, Anokye N, et al. A systematic review and economic evaluation of exercise referral schemes in primary care: a short report. Health Technol Assess 2015 Jul;19(60):1-110 [FREE Full text] [doi: 10.3310/hta19600] [Medline: 26222987]
9. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
10. Schneider PL, Bassett DR, Thompson DL, Pronk NP, Bielak KM. Effects of a 10,000 steps per day goal in overweight adults. Am J Health Promot 2006;21(2):85-89. [doi: 10.4278/0890-1171-21.2.85] [Medline: 17152246]
11. Martínez-Pérez B, de la Torre-Diez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
12. Kim H, Xie B. Health literacy in the eHealth era: a systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
13. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
14. Hills AP, Mokhtar N, Byrne NM. Assessment of physical activity and energy expenditure: an overview of objective measures. Front Nutr 2014;1:5 [FREE Full text] [doi: 10.3389/fnut.2014.00005] [Medline: 25988109]
15. Dencker M, Andersen LB. Health-related aspects of objectively measured daily physical activity in children. Clin Physiol Funct Imaging 2008 May;28(3):133-144. [doi: 10.1111/j.1475-097X.2008.00788.x] [Medline: 18241209]
16. Corder K, van Sluijs EMF, Wright A, Whincup P, Wareham NJ, Ekelund U. Is it possible to assess free-living physical activity and energy expenditure in young people by self-report? Am J Clin Nutr 2009 Mar;89(3):862-870. [doi: 10.3945/ajcn.2008.26739] [Medline: 19144732]
17. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. JAMA 2007 Nov 21;298(19):2296-2304. [doi: 10.1001/jama.298.19.2296] [Medline: 18029834]
18. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
19. Mansi S, Milosavljevic S, Baxter GD, Tumilty S, Hendrick P. A systematic review of studies using pedometers as an intervention for musculoskeletal diseases. BMC Musculoskelet Disord 2014;15:231 [FREE Full text] [doi: 10.1186/1471-2474-15-231] [Medline: 25012720]
20. Glynn LG, Hayes PS, Casey M, Glynn F, Alvarez-Iglesias A, Newell J, et al. Effectiveness of a smartphone application to promote physical activity in primary care: the SMART MOVE randomised controlled trial. Br J Gen Pract 2014 Jul;64(624):e384-e391. [doi: 10.3399/bjgp14X680461] [Medline: 24982490]
21. Hernández-Reyes A, Molina-Recio G, Molina-Luque R, Romero-Saldaña M, Cámara-Martos F, Moreno-Rojas R. Effectiveness of PUSH notifications from a mobile app for improving the body composition of overweight or obese women:

a protocol of a three-armed randomized controlled trial. BMC Med Inform Decis Mak 2020 Feb 24;20(1):40 [FREE Full text] [doi: 10.1186/s12911-020-1058-7] [Medline: 32093701]
22. Harris A, Benedict FG. Carnegie Institution of Washington. 1919. A Biometric Study of Basal Metabolism in Man URL: https://archive.org/details/biometricstudyof00harruoft/page/26/mode/2up [accessed 2020-04-13]
23. Segal KR, Edano A, Abalos A, Albu J, Blando L, Tomas MB, et al. Effect of exercise training on insulin sensitivity and glucose metabolism in lean, obese, and diabetic men. J Appl Physiol (1985) 1991 Dec;71(6):2402-2411. [doi: 10.1152/jappl.1991.71.6.2402] [Medline: 1778939]
24. Dietowin® 8.0. URL: http://dietowin.es [accessed 2020-04-04]
25. Schubert MM, Seay RF, Spain KK, Clarke HE, Taylor JK. Reliability and validity of various laboratory methods of body composition assessment in young adults. Clin Physiol Funct Imaging 2019 Mar;39(2):150-159. [doi: 10.1111/cpf.12550] [Medline: 30325573]
26. Callaway C, Chumlea W, Bouchard C, Himes J, Lohman T, Martin A. Circumferences. In: Lohman TG, Roche AF, editors. Anthropometric Standardization Reference Manual. Champaign: Human KineticsBooks; 1991:44-45.
27. Cayir Y, Aslan SM, Akturk Z. The effect of pedometer use on physical activity and body weight in obese women. Eur J Sport Sci 2015;15(4):351-356. [doi: 10.1080/17461391.2014.940558] [Medline: 25068676]
28. Turner-McGrievy GM, Wilcox S, Boutté A, Hutto BE, Singletary C, Muth ER, et al. The dietary intervention to enhance tracking with mobile devices (DIET Mobile) study: a 6-month randomized weight loss trial. Obesity (Silver Spring) 2017 Dec;25(8):1336-1342 [FREE Full text] [doi: 10.1002/oby.21889] [Medline: 28600833]
29. Richardson CR, Brown BB, Foley S, Dial KS, Lowery JC. Feasibility of adding enhanced pedometer feedback to nutritional counseling for weight loss. J Med Internet Res 2005;7(5):e56 [FREE Full text] [doi: 10.2196/jmir.7.5.e56] [Medline: 16403720]
30. Tremblay A, Therrien F. Physical activity and body functionality: implications for obesity prevention and treatment. Can J Physiol Pharmacol 2006 Feb;84(2):149-156. [doi: 10.1139/y05-132] [Medline: 16900940]
31. Slentz CA, Duscha BD, Johnson JL, Ketchum K, Aiken LB, Samsa GP, et al. Effects of the amount of exercise on body weight, body composition, and measures of central obesity: STRRIDE—a randomized controlled study. Arch Intern Med 2004 Jan 12;164(1):31-39. [doi: 10.1001/archinte.164.1.31] [Medline: 14718319]
32. Pillay JD, Kolbe-Alexander TL, van Mechelen W, Lambert EV. Steps that count: the association between the number and intensity of steps accumulated and fitness and health measures. J Phys Act Health 2014 Jan;11(1):10-17. [doi: 10.1123/jpah.2011-0288] [Medline: 23249564]
33. Bailey BW, Borup P, Tucker L, LeCheminant J, Allen M, Hebbert W. Steps measured by pedometry and the relationship to adiposity in college women. J Phys Act Health 2014 Aug;11(6):1225-1232. [doi: 10.1123/jpah.2012-0255] [Medline: 23963619]
34. Thompson DL, Rakow J, Perdue SM. Relationship between accumulated walking and body composition in middle-aged women. Med Sci Sports Exerc 2004 May;36(5):911-914. [doi: 10.1249/01.mss.0000126787.14165.b3] [Medline: 15126729]
35. Krumm EM, Dessieux OL, Andrews P, Thompson DL. The relationship between daily steps and body composition in postmenopausal women. J Womens Health (Larchmt) 2006 Mar;15(2):202-210. [doi: 10.1089/jwh.2006.15.202] [Medline: 16536684]
36. Tudor-Locke C, Bassett DR, Rutherford WJ, Ainsworth BE, Chan CB, Croteau K, et al. BMI-referenced cut points for pedometer-determined steps per day in adults. J Phys Act Health 2008;5 Suppl 1:S126-S139 [FREE Full text] [doi: 10.1123/jpah.5.s1.s126] [Medline: 18364517]
37. Hernández-Reyes A, Cámara-Martos F, Molina-Luque R, Romero-Saldaña M, Molina-Recio G, Moreno-Rojas R. Changes in body composition with a hypocaloric diet combined with sedentary, moderate and high-intense physical activity: a randomized controlled trial. BMC Womens Health 2019 Dec 27;19(1):167 [FREE Full text] [doi: 10.1186/s12905-019-0864-5] [Medline: 31882009]
38. Oikonomou EK, Antoniades C. The role of adipose tissue in cardiovascular health and disease. Nat Rev Cardiol 2019 Feb;16(2):83-99. [doi: 10.1038/s41569-018-0097-6] [Medline: 30287946]
39. Goossens GH. The metabolic phenotype in obesity: fat mass, body fat distribution, and adipose tissue function. Obes Facts 2017;10(3):207-215 [FREE Full text] [doi: 10.1159/000471488] [Medline: 28564650]
40. Karpe F, Pinnick KE. Biology of upper-body and lower-body adipose tissue—link to whole-body phenotypes. Nat Rev Endocrinol 2015 Feb;11(2):90-100. [doi: 10.1038/nrendo.2014.185] [Medline: 25365922]
41. Kang M, Marshall SJ, Barreira TV, Lee J. Effect of pedometer-based physical activity interventions: a meta-analysis. Res Q Exerc Sport 2009 Sep;80(3):648-655. [doi: 10.1080/02701367.2009.10599604] [Medline: 19791652]
42. Hargreaves EA, Mutrie N, Fleming JD. A web-based intervention to encourage walking (StepWise): pilot randomized controlled trial. JMIR Res Protoc 2016 Jan 25;5(1):e14 [FREE Full text] [doi: 10.2196/resprot.4288] [Medline: 26810251]
43. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Dec 28;14(1):42 [FREE Full text] [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
44. Conner MT, Norman P. Predicting and Changing Health Behaviour: Research and Practice with Social Cognition Models. 3rd Edition. Maidenhead: Open UniversityPress; 2015.

45. Eynon M, Foad J, Downey J, Bowmer Y, Mills H. Assessing the psychosocial factors associated with adherence to exercise referral schemes: a systematic review. Scand J Med Sci Sports 2019 May;29(5):638-650. [doi: 10.1111/sms.13403] [Medline: 30742334]
46. Rutten GM, Meis JJM, Hendriks MRC, Hamers FJM, Veenhof C, Kremers SPJ. The contribution of lifestyle coaching of overweight patients in primary care to more autonomous motivation for physical activity and healthy dietary behaviour: results of a longitudinal study. Int J Behav Nutr Phys Act 2014 Jul 16;11:86 [FREE Full text] [doi: 10.1186/s12966-014-0086-z] [Medline: 25027848]
47. Lyzwinski LN. A systematic review and meta-analysis of mobile devices and weight loss with an intervention content analysis. J Pers Med 2014;4(3):311-385 [FREE Full text] [doi: 10.3390/jpm4030311] [Medline: 25563356]

## Abbreviations

**BMI:** body mass index
**CG:** control group
**IG:** intervention group
**mHealth:** mobile health
**PA:** physical activity

*Edited by G Eysenbach; submitted 11.11.19; peer-reviewed by J Di, C Carrion; comments to author 25.01.20; revised version received 02.02.20; accepted 10.02.20; published 27.05.20.*

*Please cite as:*
*Hernández-Reyes A, Cámara-Martos F, Molina-Luque R, Moreno-Rojas R*
*Effect of an mHealth Intervention Using a Pedometer App With Full In-Person Counseling on Body Composition of Overweight Adults: Randomized Controlled Weight Loss Trial*
*JMIR Mhealth Uhealth 2020;8(5):e16999*
*URL: http://mhealth.jmir.org/2020/5/e16999/*
*doi:10.2196/16999*
*PMID:32348263*

Original Paper

# Preparing African American Men to Make Informed Prostate Cancer Screening Decisions: Development and Pilot Testing of an Interactive Online Decision Aid

Jennifer Dacey Allen[1*], MPH, SCD; Amanda Reich[1*], PhD, MPH; Adolfo G Cuevas[1], PhD; Keren Ladin[1], PhD

Department of Community Health, Tufts University, Medford, MA, United States

[*]these authors contributed equally

**Corresponding Author:**
Jennifer Dacey Allen, MPH, SCD
Department of Community Health
Tufts University
574 Boston Ave
Suite 208
Medford, MA, 02421
United States
Phone: 1 617 627 3233
Email: jennifer.allen@tufts.edu

## Abstract

**Background:** African American men are at a higher risk of developing and dying from prostate cancer compared to white men. The serum prostate-specific antigen (PSA) screening test has a high risk of false-positive results and overdiagnosis; therefore, it is not routinely recommended. Rather, men are encouraged to make individualized decisions with their medical providers, after being fully informed about its potential benefits, limitations, and risks.

**Objective:** This study aimed to describe the development and pilot testing of an interactive Web-based decision aid (DA; Prostate Cancer Screening Preparation [PCSPrep]) for African American men, designed to promote informed decision making for prostate cancer screening.

**Methods:** Four focus groups (n=33) were conducted to assess men's reactions to DAs developed in prior studies and gather information to modify the content and format. The pilot test employed a pre-posttest evaluation design. A convenience sample of 41 men aged 45-70 years with no history of prostate cancer was recruited from community settings. Participants completed online surveys before and after using PCSPrep that assessed prostate cancer screening knowledge, decision self-efficacy, decisional conflict, and preparation for decision making.

**Results:** Use of PCSPrep was associated with a significant increase in prostate cancer knowledge (49% vs 62% correct responses; $P<.001$), and men also experienced less decisional conflict (24 vs 15 on a scale of 0-100; $P=.008$). No changes in self-efficacy about decision making or screening preferences were observed. Most men (81%) reported that using PCSPrep prepared them to make informed decisions in partnership with their provider.

**Conclusions:** PCSPrep was an acceptable DA that improved men's knowledge, reduced decisional conflict, and promoted the perception of being prepared for shared decision making. Further research is needed to test the DA in a larger randomized trial.

***(JMIR Mhealth Uhealth 2020;8(5):e15502)*** doi:10.2196/15502

**KEYWORDS**

decision support techniques; prostate neoplasms; early detection of cancer; decision making (shared); men's health; minority health

## Introduction

### Background

Prostate cancer is the most common (noncutaneous) cancer among men in the United States [1]. Screening with the serum prostate-specific antigen (PSA) test is the main early detection method; it is widely utilized [2] yet remains controversial [3]. At the current time, most medical organizations advise against routine screening because of the potential risks of false-positive test results, overdiagnosis, and overtreatment. Instead, guidelines emphasize the importance of educating men about the potential risks, benefits, and harms of screening and engaging them in a process of shared decision making (SDM) with their providers [4-8]. SDM involves a discussion between the patient and the provider about the best available evidence as well as the consideration of patient preferences and values related to the screening decision [9]. However, national studies show that many patients do not experience SDM in the context of prostate cancer screening decisions [10,11]. SDM is often difficult to accomplish in clinical practice, given the short duration of clinical visits, the need to address competing health priorities, and the communication challenges between patients and providers [12].

For African American men, decision making about prostate cancer screening is particularly complex, as they are 60% more likely to be diagnosed with and nearly 2.5 times more likely to die from prostate cancer compared with white men [13]. African American men are also diagnosed at younger ages and in later stages of disease. No race-based screening guidelines are currently endorsed, and most prostate cancer screening studies have been conducted primarily among white men [14]. However, black race is acknowledged as a significant risk factor for the disease [4]. In light of this, some guidelines recommend that prostate cancer screening be made available to African American men between the ages of 45 to 50 years, but only after they are fully informed about screening [5,7].

Decision aids (DAs) are a promising means to prepare men to engage in SDM and can be administered before a clinical visit. These educational tools provide accurate and unbiased information to inform patients about the potential outcomes of a decision. DAs have generally been found to be effective in increasing patients' knowledge regarding the disease entity and screening test in question, promoting patients' sense of self-efficacy with regard to participating in SDM across a wide variety of medical decisions [15]. Over the past decade, there has been a proliferation of DAs for prostate cancer screening [16]. Taken together, these studies demonstrate that DAs are associated with increased knowledge, increased confidence in the ability to engage in decision making and reduced decisional conflict. The majority of these studies were conducted among white men in clinical settings [16].

### Objective

There is a need for DAs for African American men that incorporate disease risk assessment and can be administered in community settings or made available online [17]. This paper describes the development and pilot testing of an interactive, individually tailored Web-based DA designed specifically for African American men (Prostate Cancer Screening Preparation [PCSPrep]).

## Methods

### Conceptual Framework

The Ottawa Decision Support Framework (ODSF) [18], a mid-level theory that guides many decision support studies, provided the framework for the development and evaluation of our DA. The ODSF integrates tenets of multiple theories, including social cognitive theory, social psychology, and decision support to specify modifiable factors that can improve decision making [19]. It addresses an individual's decision needs (eg, knowledge) and articulates the components of decision support (eg, values clarification) that result in high-quality decisions that are both informed and value-based. Furthermore, the ODSF suggests a step-by-step approach for the decision-making process, which includes identification of available options, acquisition of information and skills necessary for informed decision making and SDM, clarification of values relevant to the decision, and development of a plan for action [18]. See Table 1 for examples of ODSF constructs and how they were addressed in our DA.

**Table 1.** Sample of Ottawa Decision Support Framework constructs, content, and format addressed in Prostate Cancer Screening Preparation decision aid.

| Construct | Prostate Cancer Screening Preparation content | Format/presentation |
|---|---|---|
| Knowledge | • Factual information about prostate cancer incidence and mortality among African American men; potential benefits, risks, and harms of PSA[a] screening; methods for diagnosing and treating prostate cancer, etc.<br>• Users are provided access to a section that assesses individual risk based on the risk factor information input by the user. | • Video: Doctors presenting information modeled after a popular television show. Fictionalized audience members "call in" to pose questions, which are subsequently answered by the doctors.<br>• Thermometer indicates risk relative to other men of the same age (ie, greater than, similar to, or less than average). |
| Decision self-efficacy | • Users are led through decision-making steps specified by the Ottawa Decision Support Framework, including identifying options, addressing information needs, and clarifying values. | • On-screen text with one page per step. |
| Clarification of values | • Consideration of the potential advantages and disadvantages of PSA screening. | • Users presented with common "pros" and "cons," which are rated by users. |

[a]PSA: prostate-specific antigen.

## Decision Aid Development Process

The DA development process followed steps and criteria set forth by the International Patient Decision Aid Standards (IPDAS) Collaboration. IPDAS is an international body that offers guidance to enhance the quality and effectiveness of DAs. Guidelines recommend that DAs be (1) developed with feedback from the target audience, (2) provide unbiased and detailed information in lay terms, (3) elicit information about patient needs, and (4) offer structured guidance for deliberation and communication [20]. Steps in our DA development process are described in the sections below.

### *Step 1: Obtain Feedback From the Intended Audience*

We conducted 4 focus groups with African American men (n=33) recruited from community settings (eg, churches and barbershops) through fliers and word of mouth (data not shown). Groups were facilitated by a trained African American male moderator. Focus group objectives were to assess men's reactions to DAs developed in prior studies [21], gather responses to educational messages, and assess preferences for communication strategies (eg, print, video, and online). Participants were men older than 45 years with no prior history of prostate cancer. Focus group audiotapes were professionally transcribed. We performed a thematic analysis including a hybrid of inductive and deductive approaches. First, members of the research team (JA and AR) independently reviewed each transcript and identified initial codes, based on the constructs in the interview guide. Next, team members compared codes, and through an iterative group process, they divided these codes into superordinate and subordinate categories. Following discussion and consensus, team members independently conducted line-by-line coding by compiling themes and descriptive quotes into Microsoft Excel spreadsheets. These documents were reviewed and compared. When there was a disagreement regarding the meaning of a specific quote, members of the team returned to the transcript and/or audiotape to review and come to consensus. Detailed analysis of these data is beyond the scope of this paper. Key themes included a desire for information and graphics specifically targeted for African American men, a preference for African American actors, a need for information, and skills to facilitate engagement in decision making with providers. Physicians and female family members were identified as credible information sources. Of the formats presented (print, video, and online), most men preferred an online tool that included video segments. Delivery on the internet and mobile phones was seen as acceptable among most men.

Working with a team of health communications specialists, Web developers, and experts in prostate cancer screening, we developed storyboards to test 2 potential story lines each of which used an 'edutainment' approach [22]. The first was based on a sports show with a known African American sports celebrity. The second was based on a popular network television show in which doctors talk about health issues and invite fictional audience members to ask questions. An additional 2 focus groups were conducted (k=2; n=15) to gather reactions to storyboards. Men's reactions to the television show concept were more positive than for the sports theme. Many felt that the latter was "not serious enough" and there was no need for a "big celebrity" to highlight the importance of the issue for African American men. Therefore, we proceeded with the talk show story line. The name of the DA, PCSPrep, was deemed acceptable in focus groups.

### *Step 2: Provide Unbiased and Detailed Information in Lay Terms*

The content of the DA was based on information covered in our prior DAs [21] as well as a DA developed by the Centers for Disease Control and Prevention titled "Is prostate cancer screening right for you? A decision guide for African Americans" [23], and it included a graphic from the National Cancer Institute [24]. Information was provided through video, on-screen text, and graphics. Topics addressed included the location and function of the prostate gland, incidence of prostate cancer among African American men, risk factors, methods for early detection, potential advantages and disadvantages of screening, the recommendations of major medical organizations that men make individualized decisions, meaning of an elevated

PSA test, and methods for diagnosis of prostate cancer. All information was in lay terms, and the on-screen text required only a sixth-grade reading level.

The first section of PCSPrep was a 5-min video in a talk show format hosted by 2 African American doctors (actors). Although men were able to navigate back and forth between DA segments, this video segment could not be bypassed. This was to ensure that men had all the factual information needed for an informed decision. After viewing the video, men were then able to navigate freely between any of the three remaining segments, which are described below.

### *Step 3: Elicit Information About Patient Needs*

The second segment of PCSPrep was titled "Learn More." Here, users could select from a variety of topics to gain more in-depth information (eg, risk of false positive result). The "Learn More" section also included a personalized risk assessment for prostate cancer based on the Your Disease Risk Index [25]. Users input data in response to questions about prostate cancer risk factors (eg, age, family history) and were provided with an on-screen graphic of their risk relative to other men (greater than average, average, less than average) presented as a thermometer, with higher "temperatures" indicating higher risk.

### *Step 4: Provide Structured Guidance for Deliberation and Communication*

In the third section ("Decide Now"), men were led through steps of decision making based on the ODSF. Steps included identifying decision options (screen/no screen/decide later), identifying the potential need for additional information (eg, go back to the "Learn More" section or link out to relevant medial websites), and clarifying personal values (ie, "what is most important to me?"). Values clarification included an exercise where men were presented with commonly cited pros and cons of screening (eg, "Screening could give me peace of mind" or "I don't want to have a PSA test if the results could be wrong"). They were then asked to assign a relative weight to each statement, and their responses were pictorially presented as a scale, with pros on one side and cons on the other. This information was used to guide decision making; if a man had a stated preference to undergo screening but rated the "cons" of screening more heavily than the "pros," he was told that his decision did not align with his stated values and was encouraged to revisit earlier segments of the DA to get more information and further clarify his values. The fourth segment was titled "Next Steps." This section included suggestions and tips about how to communicate one's preference and concerns to a provider as well as information about how to access screening if not otherwise available. It also included a list of questions that one might ask his provider, and a printout of this was given to users. See Figure 1 (depicting television show) and Figure 2 (step in individualized risk assessment).

**Figure 1.** The Check-Up Show.

**Figure 2.** Individualized Risk Assessment.


### Pilot Test of Prostate Cancer Screening Preparation

Men eligible to participate in the pilot test were aged 45 to 70 years, self-identified as African American, had no prior history of prostate cancer, and had not participated in focus groups. We aimed to recruit a total of 50 participants over a 3-month period. Recruitment fliers were placed in a variety of community-based organizations, including churches, barbershops, public housing, and social service agencies. Those interested in participating were screened for eligibility by phone by research assistants.

Eligible men provided written informed consent before data collection and DA use. Men completed online surveys and the DA on a study iPad in a setting that afforded privacy (eg, a meeting room). A research assistant was available to answer questions or provide assistance during survey completion and DA use. Surveys were completed immediately before and following use of the DA. The pretest took approximately 25 min to complete; the posttest took approximately 10 min. The average time spent using PCSPrep was 20 min. Men were provided with a US $50 gift card for their participation. Data collection and intervention administration took place in the summer and fall of 2015. Study procedures were approved by the Institutional Review Boards at the Dana-Farber Cancer Institute and Tufts University.

### Measures

Recognition of the PSA test was assessed with a standard item (“The prostate-specific antigen test (PSA) is a blood test that is used to find prostate cancer. Before now, had you ever heard of the PSA test?”). Subsequently, men were asked about their history of PSA testing (“Have you ever had a PSA test?” and if so, “When did you have your last PSA test?”). Assessments of the primary informed decision-making outcomes are described below. All scales with multiple survey items assessing latent constructs had good internal reliability in this sample (Cronbach alphas ranging from 0.70 to 0.91; see Table 2).

Prostate cancer knowledge was assessed with 14 questions from a validated prostate cancer knowledge scale [26]. Questions included the incidence of prostate cancer, risk factors, screening modalities as well as their limitations (false positives), diagnostic procedures, and potential treatment-related complications. The proportion of correct responses was divided by 14 to create a 0-100% scale, with higher scores indicating greater knowledge.

Decision self-efficacy, or confidence in the ability to make an informed decision and to participate in the decision making at a personally desired level, was assessed with the 11-item Decision Self-Efficacy Scale [27]. Questions ask the respondent to reflect about how confident they feel about various aspects of the decision-making process (eg, “I feel confident that I can get the facts that I need to make an informed choice”), with 3 response options including “very confident,” “somewhat confident,” and “not at all confident.” Scores are summed, divided by the total number of items, and multiplied by 25, to arrive at a range of scores from 0 (low self-efficacy) to 100 (high self-efficacy).

**Table 2.** Characteristics of study participants in the Prostate Cancer Screening Preparation pilot study (n=41).

| Characteristics | Value, n (%)[a] |
|---|---|
| **Age (years)** | |
| 45-54 | 16 (43) |
| 55-64 | 15 (41) |
| 65-70 | 6 (16) |
| **Household income (US $)** | |
| Less than $25,000 | 12 (29) |
| $25,000-$49,999 | 8 (20) |
| $50,000-$74,999 | 13 (31) |
| More than $75,000 | 8 (19) |
| **Marital status** | |
| Not married | 21 (51) |
| Married/living as married | 20 (49) |
| **Educational level** | |
| Less than high school | 2 (5) |
| Some college or 2-year degree | 10 (23) |
| 4-year college degree | 17 (41) |
| More than a 4-year college degree | 12 (29) |
| **Prior prostate-specific antigen screening** | |
| Yes | 35(85) |
| **Computer skills** | |
| Very good/good | 24 (59) |
| Fair/poor | 17 (41) |

[a]Total varies because of missing responses; percentages may not total to 100% because of rounding.

Value of screening was assessed with 8-items from our prior studies [21,28]. Participants were asked to rate their extent of agreement with statements about potential advantages (eg, "I will have peace of mind if I have prostate cancer screening") and disadvantages of screening (eg, "I do not want to have a PSA test unless doctors are reasonably certain that it can save lives") on a 4-point Likert scale (strongly agree to strongly disagree), with higher values indicating stronger agreement. Before standardizing the values scale (0-100), negative items were reverse coded such that high scores indicate a greater value placed on screening.

The Decisional Conflict Scale includes items that assess the degree to which an individual feels informed to make a decision consistent with his values, experiences uncertainty in choosing options (eg, feeling uninformed and unclear about personal values), and is likely to implement the decision. We used the low-literacy version of the scale, with 10 items, which has demonstrated good reliability and validity [29]. Questions include "Are you clear about the best choice for you?" and "Are you clear about which benefits matter most to you?" (yes/no/I don't know). Scoring is such that 0 represents no conflict and 100 reflects the highest level of conflict.

Preparedness for decision making, asked only at posttest, included a validated scale with 8 items assessing the degree to which the DA helped to provide information to make an informed decision [30]. For example, men were asked the extent to which PCSPrep helped them to *recognize* that a decision needed to be made about screening and whether they felt prepared to talk with their provider about their values related to screening. Response options were rated on a 5-point scale from "a great deal" to "not at all," with standardized higher scores (0-100) indicating greater preparedness.

We also assessed the perceived risk of prostate cancer because the DA highlighted elevated risk among African American men. This was assessed with 2 items from prior studies [31]. The first inquired about overall risk: "How likely do you think it is that you will develop prostate cancer in the next 5 years?" (very likely to very unlikely). The second asked, "Compared to the average man your age, would you say that you are...?," with response options including "more," "less," and "the same".

Decisional status was assessed by asking "If you had to decide now, what would you choose?," with response options including "I have decided to be/not to be screened" and "I don't know." Demographic characteristics, screening history, and access to health care were also assessed, using standard items from the Behavioral Risk Factor Surveillance Surveys [32].

### Analysis

Descriptive statistics were calculated to assess participants' sociodemographic characteristics and prostate cancer screening behaviors. Paired *t* tests were performed to assess changes in continuous variables (eg, knowledge and decision self-efficacy) between pretest and posttest. The McNemar test was used to assess changes in decisional status (decided or undecided) and screening preference. The Wilcoxon signed-rank test was used to assess changes in perceived risk between pretest and posttest. Cronbach alpha was used to assess the internal reliability of multi-item scales that assessed latent constructs. Effect sizes were calculated as observed average difference (delta=post minus pretest scores) divided by the standard deviation of the difference.

## Results

### Characteristics of the Study Sample

A total of 41 eligible men were recruited to the study over a 3-month period, 82% of our recruitment goal. Participants were aged between 45 and 70 years. About half the participants were married or living as married and had annual household incomes of less than US $50,000. More than two-thirds of the participants had completed a college degree or higher. A majority of the participants had heard of the PSA test and reported that they had undergone PSA screening in the past (85%). Most (78%) of the participants who had undergone PSA testing reported that they knew the results of testing, with only 7% (n=1) reporting that their PSA level had been above the normal range. About two-thirds (61%) of the participants had heard of the digital rectal exam, and 74% of the participants had undergone one in the past. Over half of the participants reported that their computer skills were "good" or "very good," and none of the participants required assistance with use of PCSPrep from the research assistant. See Table 2.

### Changes From Pretest to Posttest

Effect sizes for changes in knowledge, decisional conflict, and perceptions about the value of screening were moderate. Results are presented in Table 3. Specifically, knowledge about prostate cancer and available screening methods was very low at pretest and improved significantly after the use of PCSPrep (49% to 62%, *P*=.001; effect size=0.56). Confidence in the ability to make an informed decision (self-efficacy) was high at baseline and did not change after use of the DA (86 vs 88; *P*=.84). After using PCSPrep, men reported having lower levels of decisional conflict about screening (24 vs 15 on a scale of 0-100; *P*=.008; effect size=−0.44). Men's perceptions about the advantages of screening were high but decreased after the use of the DA (74 to 71 on a scale of 0-100; *P*=.02; effect size=−0.38). At posttest, fewer men rated their risk of developing prostate cancer to be lower than men of the same age, but this was only marginally significant (75% to 67%; *P*=.08).There was no change in the percentage of men who believed that they were "certain" or "very likely" to develop the disease in the next 5 years (data not shown). Most men (81%) reported that using PCSPrep prepared them "very well" or "well" to make informed decisions in partnership with their provided (data not shown).

**Table 3.** Changes in informed decision-making outcomes from pretest to posttest in the Prostate Cancer Screening Preparation pilot study (n=41).

| Informed decision-making outcomes | Internal reliability Cronbach alpha | Pretest, mean (SD) | Posttest, mean (SD) | Change, *P* value | Effect size |
|---|---|---|---|---|---|
| Knowledge (0-100) | .77 | 49.45 (21.52) | 61.94 (19.97) | *.001* [a] | 0.56 |
| Decision self-efficacy (0-100) | .90 | 86.12 (18.60) | 88.51 (16.89) | .84 | N/A[b] |
| Decisional Conflict Scale (*r* 0-100); Cronbach alpha=.86 | .86 | 23.8 (26.6) | 14.8 (19.52) | *.008* [a] | −0.44 |
| Value of screening (0-100) | .75 | 75.7 (13.96) | 70.67 (15.73) | *.01* [a] | −0.38 |

[a]*P*<.05.

[b]N/A: not applicable.

### Screening Decision and Preferences

There were no changes in men's screening preferences before and after using PCSPrep; at pretest, 46% of men said that they had made a definitive decision, and 47% of men reported this to be the case at posttest. The vast majority of men preferred to be screened (86%), and this did not change between test pre-test and posttest (data not shown).

## Discussion

### Principal Findings

DAs have been found to be effective interventions to complement patient/provider engagement in SDM by providing patients with information needed to assess their options and examine their values as they relate to those options. Indeed, the Centers for Medicare and Medicaid Services now requires SDM for some preference-sensitive conditions [33]. One way to achieve this is through the use of DAs. However, the development of DAs and use among African American men is incompletely understood. For African American men, the decision to undergo or forgo screening poses challenges, given their elevated risk for the disease and the controversial nature of the PSA test. It has been suggested that offering DAs for prostate cancer screening outside of a clinical setting may be particularly important for African American men who report difficulty communicating with medical providers and may have a high level of medical mistrust [34].

To our knowledge, this is the first interactive, online DA developed specifically for African American men that provides

individualized risk assessment. We found PCSPrep to be feasible to administer in community settings, even among those who reported low levels of computer skills. Moreover, men reported high levels of agreement when asked the extent to which PCSPrep helped prepare them to organize their thinking, make decisions, and have conversations about screening with their providers (ie, "preparedness for decision making").

After using PCSPrep, men had significantly greater knowledge about prostate cancer screening. However, improvements in knowledge did not translate into changes in screening preferences. Our finding of increased knowledge is consistent with other DA interventions among general audiences [35,36] and among African American men [28,37,38]. Our finding of an average 13 percentage point increase in prostate cancer knowledge is in line with these previous reports [28]. Nevertheless, having the knowledge deemed necessary to make an informed decision was still suboptimal after using the tool. Prior studies of prostate cancer screening knowledge among African American men have similarly found low levels of the knowledge required for informed decision making [28,37].

We also found that the DA did not change men's self-efficacy about making informed decisions. Men had very high levels of confidence in their ability to make informed decisions before using the DA, despite having relatively low levels of knowledge. Knowledge scores were not correlated with decision self-efficacy (Pearson $r$=−0.25; $P$=.12). We believe that further attention to the relationship between knowledge and decision self-efficacy is warranted in future studies.

After engaging with PCSPrep, men perceived fewer advantages of screening, a phenomenon that has been consistently reported [16]. Despite this, overall opinions about screening were universally favorable, with the majority valuing the benefits over the potential risks and harms. Similarly, other studies have also found that men tend to prefer prostate cancer screening even in light of its limitations. Indeed, few people are aware of the concept of over detection or can identify potential harms of screening [39,40], and few people decline screening even when provided with information about risks of false-positive test results [41,42].

We found that men had less decisional conflict after using PCSPrep, which is also consistent with prior DA studies [35,36]. Presumably, improvements in knowledge and clarification of preferences reduces men's ambivalence about decision making and improves decision quality. A study using structural equation modeling found that increased knowledge after DA use had an indirect effect on decisional conflict by increasing the perceived risk and decreasing anxiety about decision making [43]. In our study, men gained significant knowledge and were more likely to perceive themselves to be at a higher risk compared with men of the same age. At the same time, decisional conflict was reduced. However, we did not find knowledge to be significantly correlated with decisional conflict at posttest (Pearson $r$=−0.24; $P$=.13). Future research should examine other intraindividual factors, such as cultural barriers and decision-making preferences, that may play a role in decisional conflict.

### Strengths and Limitations

This study has limitations that must be acknowledged. First, we used a quasi-experimental design with a small convenience sample with no control group. We cannot rule out alternative explanations for findings nor can we generalize these results to the broader population of African American men. Moreover, we did not assess actual screening behaviors, men's actual discussions with their providers, or long-term retention of knowledge or skills. These are limitations in the existing literature, and these longer-term outcomes warrant further study. We also acknowledge that men in this sample were more highly educated than the general US population [44], with two-thirds of the men having completed at least a bachelor's degree. Although we designed the DA for a population with a sixth-grade reading level, we cannot assume that it would have the same impact in a sample with lower levels of education.

Despite these limitations, we believe that this study can offer some new insights for prostate cancer screening DAs. This is among the first DAs designed for African American men that integrate personalized risk estimates and other interactive functions, including values clarification exercises. We also specifically addressed the issue of false-positive test results in the DA, which has been called for in two recent systematic reviews of prostate screening DAs [35,36]. These features may enhance attention to messages, increase understanding and recall, and potentially lead to improved quality of decisions. If the DA is ultimately found efficacious in a larger randomized controlled trial, the online format has the potential to reach broad, geographically dispersed populations.

Moreover, this paper is responsive to the recent calls for the explicit articulation of theoretical underpinnings for DA interventions [45]. The application and testing of theoretical models could enhance the understanding of the mechanisms through which DAs operate and ultimately improve their efficacy. We observed that fewer men believed that their risk of developing prostate cancer compared with men in the same age group increased but their perceived risk of developing the disease in the next 5 years was unchanged. This suggests that decision support may improve the accuracy of disease risk among this population, and at the same time, enable men to make decisions without undue internal conflict. Future research is needed to determine if existing conceptual models for decision support interventions, which tend to emphasize knowledge acquisition, can have a greater impact on other decision-making outcomes, including actual screening decisions and engagement in SDM with providers.

## Conflicts of Interest

None declared.

## References


1. American Cancer Society. Cancer Facts and Statistics URL: https://www.cancer.org/research/cancer-facts-statistics.html [accessed 2019-03-04]
2. Leyva B, Persoskie A, Ottenbacher A, Hamilton JG, Allen J, Kobrin S, et al. Do men receive information required for shared decision making about PSA testing? Results from a national survey. J Cancer Educ 2016 Dec;31(4):693-701 [FREE Full text] [doi: 10.1007/s13187-015-0870-8] [Medline: 26498649]
3. Kim EH, Andriole GL. Prostate-specific antigen-based screening: controversy and guidelines. BMC Med 2015 Mar 24;13:61 [FREE Full text] [doi: 10.1186/s12916-015-0296-5] [Medline: 25857320]
4. American Academy of Family Physicians. Clinical Preventive Service Recommendation: Prostate Cancer URL: https://www.aafp.org/patient-care/clinical-recommendations/all/prostate-cancer.html [accessed 2019-03-04]
5. Carter HB. American Urological Association (AUA) guideline on prostate cancer detection: process and rationale. BJU Int 2013 Sep;112(5):543-547 [FREE Full text] [doi: 10.1111/bju.12318] [Medline: 23924423]
6. Chou R, Croswell J, Dana T, Bougatsos C, Blazina I, Fu R, et al. Screening for prostate cancer: a review of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2011 Dec 6;155(11):762-771. [doi: 10.7326/0003-4819-155-11-201112060-00375] [Medline: 21984740]
7. Moyer VA, US Preventive Services Task Force. Screening for prostate cancer: US Preventive Services Task Force recommendation statement. Ann Intern Med 2012 Jul 17;157(2):120-134. [doi: 10.7326/0003-4819-157-2-201207170-00459] [Medline: 22801674]
8. Qaseem A, Barry M, Denberg T, Owens D, Shekelle P, Clinical Guidelines Committee of the American College of Physicians. Screening for prostate cancer: a guidance statement from the Clinical Guidelines Committee of the American College of Physicians. Ann Intern Med 2013 May 21;158(10):761-769. [doi: 10.7326/0003-4819-158-10-201305210-00633] [Medline: 23567643]
9. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
10. Fedewa SA, Gansler T, Smith R, Sauer A, Wender R, Brawley O, et al. Recent patterns in shared decision making for prostate-specific antigen testing in the United States. Ann Fam Med 2018 Mar;16(2):139-144 [FREE Full text] [doi: 10.1370/afm.2200] [Medline: 29531105]
11. Han PK, Kobrin S, Breen N, Joseph D, Li J, Frosch D, et al. National evidence on the use of shared decision making in prostate-specific antigen screening. Ann Fam Med 2013;11(4):306-314 [FREE Full text] [doi: 10.1370/afm.1539] [Medline: 23835816]
12. Tan AS, Mazor KM, McDonald D, Lee SJ, McNeal D, Matlock DD, et al. Designing shared decision-making interventions for dissemination and sustainment: can implementation science help translate shared decision making into routine practice? MDM Policy Pract 2018;3(2):2381468318808503 [FREE Full text] [doi: 10.1177/2381468318808503] [Medline: 30560204]
13. Chornokur G, Dalton K, Borysova M, Kumar N. Disparities at presentation, diagnosis, treatment, and survival in African American men, affected by prostate cancer. Prostate 2011 Jun 15;71(9):985-997 [FREE Full text] [doi: 10.1002/pros.21314] [Medline: 21541975]
14. Shenoy D, Packianathan S, Chen AM, Vijayakumar S. Do African-American men need separate prostate cancer screening guidelines? BMC Urol 2016 May 10;16(1):19 [FREE Full text] [doi: 10.1186/s12894-016-0137-7] [Medline: 27165293]
15. Stacey D, Légaré F, Lewis K, Barry M, Bennett C, Eden K, et al. Decision aids for people facing health treatment or screening decisions. Cochrane Database Syst Rev 2017 Apr 12;4:CD001431 [FREE Full text] [doi: 10.1002/14651858.CD001431.pub5] [Medline: 28402085]
16. Ivlev I, Jerabkova S, Mishra M, Cook L, Eden K. Prostate cancer screening patient decision aids: A systematic review and meta-analysis. Am J Prev Med 2018 Dec;55(6):896-907 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.016] [Medline: 30337235]
17. Jimbo M, Rana GK, Hawley S, Holmes-Rovner M, Kelly-Blake K, Nease DE, et al. What is lacking in current decision aids on cancer screening? CA Cancer J Clin 2013 May;63(3):193-214 [FREE Full text] [doi: 10.3322/caac.21180] [Medline: 23504675]
18. Légaré F, O'Connor AC, Graham I, Saucier D, Côté L, Cauchon M, et al. Supporting patients facing difficult health care decisions: use of the Ottawa Decision Support Framework. Can Fam Physician 2006 Apr;52:476-477 [FREE Full text] [Medline: 17327891]
19. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. Science 1981 Jan 30;211(4481):453-458. [doi: 10.1126/science.7455683] [Medline: 7455683]
20. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International Patient Decision Aids Standards (IPDAS) Collaboration. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. Br Med J 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
21. Allen JD, Othus M, Hart A, Tom L, Li Y, Berry D, et al. A randomized trial of a computer-tailored decision aid to improve prostate cancer screening decisions: results from the Take the Wheel trial. Cancer Epidemiol Biomarkers Prev 2010 Sep;19(9):2172-2186 [FREE Full text] [doi: 10.1158/1055-9965.EPI-09-0410] [Medline: 20716619]

22. Volk RJ, Jibaja-Weiss M, Hawley S, Kneuper S, Spann S, Miles B, et al. Entertainment education for prostate cancer screening: a randomized trial among primary care patients with low health literacy. Patient Educ Couns 2008 Dec;73(3):482-489 [FREE Full text] [doi: 10.1016/j.pec.2008.07.033] [Medline: 18760888]
23. United States Department of Health and Human Services, Centers for Disease Control and Prevention. Us Too.: US Department of Health and Human Services Prostate Cancer Screening: A Decision Guide for African Americans URL: https://www.ustoo.org/PDFs/CDC_PCa_Screen_Guide_AA.pdf [accessed 2015-08-05]
24. National Cancer Institute. Prostate Cancer Screening (PDQ)–Patient Version URL: https://www.cancer.gov/types/prostate/patient/prostate-screening-pdq [accessed 2019-11-21]
25. Colditz CG, Atwood K, Emmons K, Monson R, Willett W, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard Cancer Risk Index. Risk Index Working Group, Harvard Center for Cancer Prevention. Cancer Causes Control 2000 Jul;11(6):477-488. [doi: 10.1023/a:1008984432272] [Medline: 10880030]
26. Radosevich DM, Partin M, Nugent S, Nelson D, Flood A, Holtzman J, et al. Measuring patient knowledge of the risks and benefits of prostate cancer screening. Patient Educ Couns 2004 Aug;54(2):143-152. [doi: 10.1016/S0738-3991(03)00207-6] [Medline: 15288907]
27. Bunn H, O'Connor A. Validation of client decision-making instruments in the context of psychiatry. Can J Nurs Res 1996;28(3):13-27. [Medline: 8997937]
28. Allen JD, Mohllajee A, Shelton R, Drake B, Mars D. A computer-tailored intervention to promote informed decision making for prostate cancer screening among African American men. Am J Mens Health 2009 Dec;3(4):340-351 [FREE Full text] [doi: 10.1177/1557988308325460] [Medline: 19477736]
29. O'Connor AM, Jacobsen M, Stacey D. An evidence-based approach to managing women's decisional conflict. J Obstet Gynecol Neonatal Nurs 2002;31(5):570-581. [doi: 10.1111/j.1552-6909.2002.tb00083.x] [Medline: 12353737]
30. Bennett C, Graham I, Kristjansson E, Kearing S, Clay K, O'Connor AM. Validation of a preparation for decision making scale. Patient Educ Couns 2010 Jan;78(1):130-133. [doi: 10.1016/j.pec.2009.05.012] [Medline: 19560303]
31. Weinstein ND, Kwitel A, McCaul K, Magnan R, Gerrard M, Gibbons F. Risk perceptions: assessment and relationship to influenza vaccination. Health Psychol 2007 Mar;26(2):146-151. [doi: 10.1037/0278-6133.26.2.146] [Medline: 17385965]
32. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System Questionnaires URL: https://www.cdc.gov/brfss/questionnaires/index.htm [accessed 2019-07-10]
33. Centers for Medicare and Medicaid. CMS Quality Strategy 2016 URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Downloads/CMS-Quality-Strategy.pdf [accessed 2019-09-15]
34. Allen JD, Kennedy M, Wilson-Glover A, Gilligan T. African-American men's perceptions about prostate cancer: implications for designing educational interventions. Soc Sci Med 2007 Jun;64(11):2189-2200. [doi: 10.1016/j.socscimed.2007.01.007] [Medline: 17399877]
35. Ilic D, Jammal W, Chiarelli P, Gardiner R, Hughes S, Stefanovic D, et al. Assessing the effectiveness of decision aids for decision making in prostate cancer testing: a systematic review. Psychooncology 2015 Oct;24(10):1303-1315. [doi: 10.1002/pon.3815] [Medline: 25873433]
36. Riikonen RJ, Guyatt G, Kilpeläinen TP, Craigie S, Agarwal A, Agoritsas T, et al. Decision aids for prostate cancer screening choice: a systematic review and meta-analysis. JAMA Intern Med 2019 Jun 24 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0763] [Medline: 31233091]
37. Ellison GL, Weinrich SP, Lou M, Xu H, Powell IJ, Baquet CR. A randomized trial comparing web-based decision aids on prostate cancer knowledge for African-American men. J Natl Med Assoc 2008 Oct;100(10):1139-1145 [FREE Full text] [doi: 10.1016/s0027-9684(15)31481-4] [Medline: 18942274]
38. Owens OL, Felder T, Tavakoli A, Revels A, Friedman D, Hughes-Halbert C, et al. Evaluation of a computer-based decision aid for promoting informed prostate cancer screening decisions among African American men: iDecide. Am J Health Promot 2019 Feb;33(2):267-278. [doi: 10.1177/0890117118786866] [Medline: 29996666]
39. van den Bruel A, Jones C, Yang Y, Oke J, Hewitson P. People's willingness to accept overdetection in cancer screening: population survey. Br Med J 2015 Mar 3;350:h980 [FREE Full text] [doi: 10.1136/bmj.h980] [Medline: 25736617]
40. Sutkowi-Hemstreet A, Vu M, Harris R, Brewer N, Dolor R, Sheridan S. Adult patients' perspectives on the benefits and harms of overused screening tests: a qualitative study. J Gen Intern Med 2015 Nov;30(11):1618-1626 [FREE Full text] [doi: 10.1007/s11606-015-3283-9] [Medline: 25869017]
41. de Bekker-Grob EW, Rose J, Donkers B, Essink-Bot M, Bangma C, Steyerberg E. Men's preferences for prostate cancer screening: a discrete choice experiment. Br J Cancer 2013 Feb 19;108(3):533-541 [FREE Full text] [doi: 10.1038/bjc.2013.5] [Medline: 23361056]
42. Perneger TV, Cullati S, Schiesari L, Charvet-Bérard A. Impact of information about risks and benefits of cancer screening on intended participation. Eur J Cancer 2010 Aug;46(12):2267-2274. [doi: 10.1016/j.ejca.2010.04.011] [Medline: 20466537]
43. Stephens RL, Xu Y, Volk R, Scholl L, Kamin S, Holden E, et al. Influence of a patient decision aid on decisional conflict related to PSA testing: a structural equation model. Health Psychol 2008 Nov;27(6):711-721. [doi: 10.1037/0278-6133.27.6.711] [Medline: 19025266]

44. de Brey C, Musu L, McFarland J, Wilkinson-Flicker S, Diliberti M, Zhang A. National Center for Education Statistics. 2019 Feb. Status and Trends in the Education of Racial and Ethnic Groups 2018 URL: https://nces.ed.gov/pubs2019/2019038.pdf [accessed 2020-02-19]
45. Durand MA, Stiel M, Boivin J, Elwyn G. Where is the theory? Evaluating the theoretical frameworks described in decision support technologies. Patient Educ Couns 2008 Apr;71(1):125-135. [doi: 10.1016/j.pec.2007.12.004] [Medline: 18242040]

## Abbreviations

**DA:** decision aid
**IPDAS:** International Patient Decision Aid Standards
**ODSF:** Ottawa Decision Support Framework
**PCSPrep:** Prostate Cancer Screening Preparation
**PSA:** prostate-specific antigen
**SDM:** shared decision making

*Edited by G Eysenbach; submitted 15.07.19; peer-reviewed by C Dorfman, M Dogba, L Forbes; comments to author 08.10.19; revised version received 03.12.19; accepted 24.01.20; published 05.05.20.*

*Please cite as:*
*Allen JD, Reich A, Cuevas AG, Ladin K*
*Preparing African American Men to Make Informed Prostate Cancer Screening Decisions: Development and Pilot Testing of an Interactive Online Decision Aid*
*JMIR Mhealth Uhealth 2020;8(5):e15502*
*URL: https://mhealth.jmir.org/2020/5/e15502*
*doi:10.2196/15502*
*PMID:32369032*

Original Paper

# Effectiveness of Text Message Interventions for Weight Management in Adolescents: Systematic Review

Stephanie Ruth Partridge[1,2], BSc (Hons I), PhD, APD; Rebecca Raeside[1], B Bio Med Sc, MPH; Anna Singleton[1], BSc (Hons 1), MSc; Karice Hyun[1], BSc, MAppStat, PhD; Julie Redfern[1,3], BSc, BAppSc (Physio Hons 1), PhD

[1]The University of Sydney, Westmead Applied Research Centre, Faculty of Medicine and Health, Sydney, Australia

[2]The University of Sydney, Prevention Research Collaboration, Charles Perkins Centre, Sydney School Public Health, Sydney, Australia

[3]The George Institute for Global Health, Sydney, Australia

**Corresponding Author:**
Stephanie Ruth Partridge, BSc (Hons I), PhD, APD
The University of Sydney
Westmead Applied Research Centre
Faculty of Medicine and Health
Department of General Practice
Acacia House, Westmead Hospital
Sydney, 2145
Australia
Phone: 61 288908187
Email: stephanie.partridge@sydney.edu.au

## Abstract

**Background:** The incidence of obesity among adolescents is increasing. Text messages are a primary communication form for adolescents and potentially a scalable strategy for delivering population health interventions.

**Objective:** This study aimed to determine the effectiveness of text message interventions in reducing BMI in adolescents and describe characteristics that are common to effective interventions.

**Methods:** This systematic review included randomized controlled trials of text message lifestyle interventions involving adolescents aged 10 to 19 years with outcomes focused on obesity prevention or management. Primary outcome was objective or self-report change in BMI.

**Results:** In total, 4362 records were identified, and 215 full-text articles were assessed for eligibility. A total of 8 unique studies were identified, including 767 participants, mean age 14.3 (SD 0.9) years, BMI 29.7 (SD 1.6) kg/m2 and 53.1% (407/767) female (31/101, 30.7%-172/172, 100.0%). All interventions were multicomponent. The median active intervention period was 4.5 months. During the active and extended intervention phases, text messages accounted for >50% (8 studies) and >85% (3 studies) of contact points, respectively. Text messages were heterogeneous, with a median of 1.5 text messages sent per week (range: 1-21). A total of 4 studies utilized two-way text message communication with health professionals Of the 8 studies, 7 demonstrated reductions in BMI or BMI z-score in the intervention group compared with the control at the end of the final follow-up. The effect was only statistically significant in 1 study at 6 months. Over 6 months, reductions in BMI (kg/m2) ranged from 1.3% to 4.5% and BMI z-score ranged from 4.2% to 28.1%. Overall quality of the studies was low.

**Conclusions:** Further research is required to elucidate the effectiveness and potential impact of text message interventions on weight and weight-related behaviors in adolescents.

***(JMIR Mhealth Uhealth 2020;8(5):e15849)*** doi:10.2196/15849

**KEYWORDS**

adolescent; text message; obesity; overweight; prevention; mHealth

## Introduction

### Background

The prevalence of overweight and obesity among children and adolescents is now estimated to be 18% globally, with the prevalence of overweight increasing by 47% over the last four decades [1]. Lifestyle risk factors, namely, suboptimal diet, physical inactivity, and overweight and obesity are well established during adolescence [2-4], making this second decade of life (10-19 years) [5] a critical period for the development of lifelong health trajectories [6]. Weight gain during adolescence is associated with earlier onset type 2 diabetes mellitus and cardiovascular disease (CVD) [2,7]. Compared with adolescents without obesity, adolescents with obesity have increased levels of circulating free fatty acids, which reduce insulin sensitivity, potentially contributing to impaired insulin secretion [7]. Moreover, adolescents who gain weight and maintain a high BMI into adulthood have higher odds of developing hypertension, dyslipidemia, and systemic inflammation [2]. More short-term adverse health outcomes associated with adolescent overweight and obesity include weight stigma [8] and reduced quality of life and self-esteem [9]. Interventions to prevent obesity, including combinations of both physical activity and diet interventions have so far remained mainly school based and with limited evidence of effectiveness on BMI [10]. Therefore, for health systems to impact population overweight and obesity, interventions need to be personalized, low cost, and scalable with a broad reach to engage all adolescents at risk of obesity.

The ubiquity of mobile phones has given rise to new ways of delivering health care through mobile health (mHealth) [11]. mHealth has the potential to provide personalized, low-cost, and population-wide behavior change programs for obesity prevention and management, particularly for adolescents who are digital frontrunners [12]. Adolescent mobile phone ownership is high in both developed and developing countries. For example, over 90% of adolescents in Australia and the United States own a mobile phone [13,14] and the average age to first own a mobile phone was 10.4 years in Korean children [15]. Moreover, adolescents spend on average more than 3 hours per day using a mobile phone, with their device always nearby [16], and text messaging remains a primary means of communication [13]. As such, text message interventions have the potential to be a wide-reaching and engaging mode of health care delivery for adolescents.

There is evidence which indicates that text messaging interventions in adults can promote improvements in lifestyle behavior change for weight loss, smoking cessation, physical activity, management of chronic disease and CVD risk, blood pressure lowering, and diabetes care [17-22]. Text messages are a feasible and acceptable form of communication for adolescents with obesity [23,24]. Features of text messages that are well suited to dietary and physical activity behavior change interventions for adolescent weight management include two-way communication between participant and health professional; direct communication with participants (ie, not required to actively log on to a smartphone app or website for support); and the ability to prompt and reward the repetition of positive behavior change in real time.

A critical step to translate and disseminate mHealth obesity prevention and management programs for adolescents is to understand the effectiveness and acceptability of the mHealth intervention components, which may help improve the integration of evidence-based practices into health systems. Previous systematic reviews have found limited studies utilizing text messages for improving diet and physical activity behaviors of adolescents [25,26]. However, Rose et al [26] did identify successful behavior change strategies for improved nutrition and physical activity behaviors, namely, education, goal setting, and self-monitoring, which are feasible for delivery via text message. Another review investigating mobile technologies for the prevention and treatment of pediatric obesity found considerable heterogeneity in the three included studies utilizing text messages [27].

### Objective

In recent years, evidence on the effectiveness of interventions delivered via text messages for adolescent weight management have been accumulating but not yet consolidated. Moreover, the collective evidence on text message development and process evaluation data, such as how adolescents engage with and use the text messages, have not been synthesized. Therefore, the aim of this systematic review was (1) to evaluate the effect of interventions delivered via text messages on weight outcomes in an adolescent population and (2) to understand intervention characteristics that are common to effective interventions delivered by text message.

## Methods

### Protocol and Registration

This systematic review was conducted and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement guidelines [28] (Multimedia Appendix 1) and followed the predetermined methods documented in a protocol. The review was registered in the International Prospective Register of Systematic Reviews (PROSPERO; registration number: CRD42018109197).

### Eligibility Criteria

Studies were included that met the following criteria: (1) randomized controlled trials (RCTs) that included lifestyle interventions of any duration which included mobile phone SMS or text message. Only RCTs were included as they provide the strongest evidence for the benefits of a health care intervention [29]; (2) participants were adolescents, defined by the World Health Organization as the second decade of life, 10 to 19 years [5], both girls and boys, not pregnant and free of acute illness or chronic disease, as some conditions may influence body weight outcomes or ability to change lifestyle behaviors; (3) only studies focused on obesity prevention or management were included; and (4) studies with interventions of any duration that involved the delivery of text messages via a mobile phone device, including multicomponent interventions that were delivered in part by text message. Studies with interventions that use at least any two of the behavioral

techniques (BCTs) to achieve behavior change, for example, goal setting, were included. BCTs are defined as a mechanism of action in an intervention that contributes to positive behavior change [30]. The mode of text message delivery included standard SMS or messaging apps like WhatsApp (studies were excluded if they included text messages that are only appointment or other reminders); (5) all study settings were included, that is, health care, community, home, or school based; (6) a comparator group of participants receiving standard care (no messages or some form of control text message); (7) a study outcome of change in body weight measured in BMI or BMI z-score (also called BMI standard deviation scores); (8) studies published in any language were considered; and (9) studies published after 2005 were considered. The cut-off date of 2005 was selected as the current generation of adolescents (*Generation Z*) appeared in the population after 1995, and the oldest of this generation were 10 years old in 1995. The criteria for included studies in this review are summarized in the Population, Intervention, Comparator, Outcomes, and Setting format in Table 1.

**Table 1.** Summary description of Population, Intervention, Comparator, Outcomes, and Setting (PICOS) components.

| PICOS components | Description |
|---|---|
| Population | Individuals (adolescents, 10-19 years) of any demographic background |
| Intervention | Interventions that include mobile phone SMS or texting intervention |
| Comparator | Intervention vs usual care |
| Outcomes | Changes in body weight measured in terms of BMI and/or BMI z-score[a] |
| Setting | Randomized controlled trials conducted in any setting |

[a]Units BMI is above or below average BMI for age- and sex-specific reference values.

## Information Sources and Searches

A total of 10 major electronic databases (Pre-Medline, MEDLINE, Cochrane, Cochrane Central Register of Controlled Trials, Excerpta Medica Database, Cumulative Index of Nursing and Allied Health Literature, Allied and Complementary Medicine Database, Informit, Scopus, and Web of Science) were systematically searched until January 21, 2019, and hand searching was conducted until July 12, 2019. The database searches were developed in collaboration with a librarian. Search terms included combinations, truncations, and synonym of the following: (1) adolescent; (2) phone text messaging (mHealth or *telemedicine* or *tele-medicine* or *mobile phone* or *mobile phones* or *cell phones* or *cell phone* or *cell phones* or SMS or *short message service* or *short message* or texting or *text message* or *text messages* or *text messaging* or *text messaged* or *text messaging* or txt or text); and (3) weight loss or weight maintenance (*diet, reducing* or *weight reduction programs* or *weight loss* or *weight maintenance* or *weight loss* or *weight management* or weight and loss or manage* or maintenance or maintain or maintain*). Appropriate RCT filters were used to maximize the identification of RCTs. Additional articles were obtained through a hand search of reference lists, conference proceedings, key journals in the field, abstracts, clinical trial databases, and by contacting experts in the field. Full electronic search strategies for each database and screenshots of all search results are available in Multimedia Appendices 2 and 3.

## Study Selection

One author (SP) carried out all electronic database searches. Search results across databases were merged using reference management software, Endnote X8 (Camelot UK Bidco Limited, Clarivate Analytics, United Kingdom), and duplicate records of the same study were removed. Study selection followed the process described in the Cochrane Handbook of Systematic Reviews and PRISMA statements. Two researchers (SP and RR) independently screened titles and abstracts to remove irrelevant studies to identify studies that met the inclusion criteria described above using a predetermined eligibility assessment form. Any disagreements were discussed and resolved by consensus between two authors (SP and RR), and a third author was consulted (AS) in the case of unresolved disagreement.

## Data Collection Process

For studies meeting the inclusion criteria, information was extracted using a predesigned electronic data extraction table based on PRISMA statement [28] and data items required for the Cochrane Collaboration's risk of bias tool [29] that was developed and test-piloted for this review. One author extracted (SP) the data, and a second author (RR) independently cross-checked a random 20% of the data for accuracy. Extracted data included data items on study characteristics (design, aim, sample size, active intervention duration, extended intervention duration, follow-up time points, attrition, comparison of dropouts, and recruitment methods), participant characteristics (age, weight, BMI ($kg/m^2$) and/or BMI z-score at baseline, gender, and ethnicity), intervention characteristics (components, theoretical underpinning, BCTs using the behavior-specific taxonomy of 40 BCTs for physical activity and healthy eating behaviors [CALO-RE taxonomy] [30], setting, peer support, personnel interaction, text message details, and comparators), study outcome measures (method of assessment, changes in body weight measured in BMI change in $kg/m^2$, and/or BMI z-score change in units during the intervention and all follow-up period[s], changes in diet, physical activity and/or psychosocial well-being if reported or collected, measures of error, and statistical significance), and adherence measures (number, type, and definition of adherence measures). When two or more articles reported results from the same study, all articles were considered together for complete data extraction. Authors were contacted for missing, incomplete, or unclear data.

### Data Synthesis and Analysis

The key characteristics of the included studies were summarized in text form and tabulated using the information collected from the data extraction form. The primary outcome of interest was the change in body weight measured in BMI change in kg/m$^2$ or BMI z-score change in units. Where possible, for all study arms, the mean or median change was recorded at baseline, postintervention, and any additional follow-up(s). Measures of error were standard error or standard deviation and associated *P* values for change between groups at follow-up(s), and overtime were recorded if available. The clinical significance of outcomes was also considered. Modest reductions in BMI z-score (0.01-0.15) in adolescents with overweight or obesity have been associated with improvements in several cardiovascular risk factors and considered to be clinically meaningful [31]. More significant improvements in cardiovascular risk can be seen with BMI z-score reductions of >0.25 in adolescents with obesity [32]. Interventions were heterogeneous; only 5 studies reported the mean change in BMI, only 4 studies reported the mean change in BMI z-score, and all were at varying follow-up time points. This heterogeneity rendered the sample size too small for a meta-analysis of these outcomes to be conducted at each follow-up time point [33].

### Risk of Bias Assessment

#### *Cochrane Risk of Bias Assessment*

The Cochrane Collaboration's tool was used to assess the risk of bias at the individual study level [29]. The primary sources of systematic bias in trials were assessed including the selection of participants (random sequence generation and allocation concealment methods); performance (blinding of participants and personnel); detection (blinding of outcome assessment); attrition (incomplete outcome data); and reporting (selective reporting of study outcomes). Two authors (SP and RR) independently evaluated each study for risk of bias and permitted a judgment of low risk, high risk, or unclear risk. A third author (AS) was consulted in the case of unresolved disagreement.

### Quality of Evidence Assessment

#### *Grading of Recommendations Assessment, Development, and Evaluation Assessment*

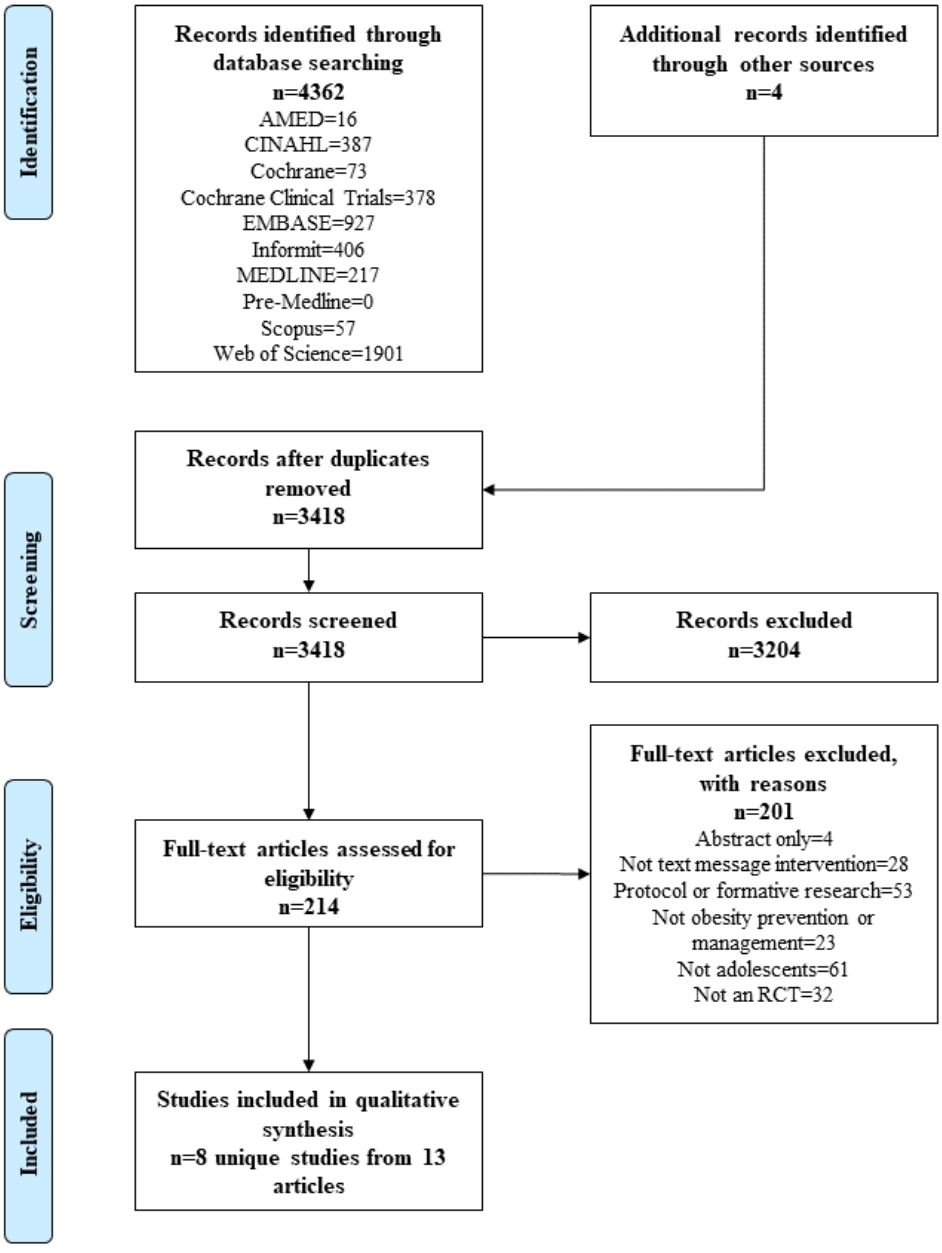
The quality of the body of evidence was determined using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system [34]. In total, 5 categories were considered to ascribe a quality rating: limitations in study designs; consistency of results; the directness of the evidence concerning study populations, intervention design, and outcomes measured; the precision of outcomes; and the presence of publication biases. Two authors (SP and RR) independently evaluated the quality of the body of evidence. A third author (AS) was consulted in the case of unresolved disagreement.

## Results

### Study Selection

The search found 4362 articles from all electronic database searches and an additional 4 articles through hand searching of reference lists (Figure 1). After exclusion of duplicates, 3418 articles were screened by title and abstract, and 3203 were excluded. A total of 215 full-text articles were assessed for edibility, and 201 were excluded with reasons outlined in Figure 1 and Multimedia Appendix 4. Eight unique studies [35-42] from 13 publications were included in this review [24,35-41,43-47].

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) flow diagram of included studies. RCT: randomized controlled trial. CINAHL: Cumulative Index of Nursing and Allied Health Literature; AMED: Allied and Complementary Medicine Database.




## Effectiveness of Interventions

### *Body Mass Index*

All studies measuring BMI (kg/m$^2$) measured height and weight in person by trained personnel using standardized procedures and calibrated equipment (Table 2). The 5 studies calculating BMI z-score specified the country of origin for the age- and sex-specific reference values [37,39-42]. Of the 8 studies, 7 demonstrated reductions in BMI or BMI z-score in the intervention group compared with the control at the end of the final follow-up (Table 3). The effect was only statistically significant in 1 study at 6 months [37]. Significant time by group effects were observed in 3 studies [36,37,40]; 2 studies reported the effect of the intervention on both BMI and BMI z-score [37,40], 3 studies reported the effect of the intervention on BMI only [35,36,38], and 3 studies reported the effect of the

intervention on BMI z-score only [39,41,42]. A decrease in BMI ranged between 1.3% and 4.5% in the intervention group compared with the control at 6 months, with the exception of 1 study, which saw an increase in BMI of 3.8% in the intervention group compared with the control group at 6 months [38]. The 3 studies which reported BMI z-score change at 6 months varied with a decrease of 28.1%, 4.2%, and 4.5% in the intervention group compared with the control [37,41,42].

**Table 2.** Method of BMI and BMI z-score assessment (n=8).

| Parameter, author, year, country | Method of assessment |
|---|---|
| **BMI** | |
| Abraham et al, 2015, China [35] | Height and weight measured in person, by a trained researcher using standard procedures and calibrated equipment |
| Bagherniya et al, 2018, Iran [36] | Height and weight measured in person by a trained researcher in duplicate using standard procedures and calibrated equipment |
| Chen et al, 2017, the United States [37] | Height and weight measured in person, by trained research assistant who was blinded to group assignment |
| Love-Osborne et al, 2016, the United States [38] | Height and weight measured in person with shoes off, repeated twice by study staff |
| Nguyen et al, 2012, Australia [40,46] | Height and weight measured in person by a trained researcher using standard procedures and calibrated equipment |
| **BMI z-score** | |
| Chen et al, 2017, the United States [37] | BMI z-scores based on age- and sex-specific reference values (US reference data) |
| Jensen et al, 2019, the United States [42] | BMI z-scores based on age- and sex-specific reference values (US reference data) |
| Mameli et al, 2018, Italy [39] | BMI z-scores based on age- and sex-specific reference values (Italian reference data) |
| Nguyen et al, 2012, Australia [40,46] | BMI z-scores based on age- and sex-specific reference values (US reference data) |
| Patrick et al, 2013, the United States [41] | BMI z-scores based on age- and sex-specific reference values (US reference data) |

**Table 3.** Mean change in BMI or BMI-z score at follow up(s) (n=8).

| Author, year, citation, outcome, follow-up timepoint (months), and study arms | Mean difference within groups (SD) | Effect size (%) | Level of effect size[a] | Mean difference between groups | Mean difference between groups over time |
|---|---|---|---|---|---|
| **Abraham et al, 2015, China [35]** | | | | | |
| **BMI** | | | | | |
| **3** | | **0.6 decrease** | **+** | ***P*=.14** | **NR[b]** |
| IT[c] | 0 (NR) | | | | |
| sLMP[d] | 0.4 (NR) | | | | |
| C[e] | 0.2 (NR) | | | | |
| **6** | | **1.3 decrease** | **+** | ***P*=.07** | **NR** |
| IT | −0.1 (NR) | | | | |
| sLMP | −0.5 (NR) | | | | |
| C | 0.3 (NR) | | | | |
| **Bagherniya et al, 2018, Iran [36]** | | | | | |
| **BMI** | | | | | |
| **3.5** | | | | | |
| I[f] | NR | NR | NR | NR | NR |
| C | NR | —[g] | — | — | — |
| **7** | | | | | |
| I | −0.7 (NR) | 3.9 decrease | + | *P*=.13 | *P*<.001 |
| C | 0.4 (NR) | — | — | — | — |
| **Chen et al, 2017, the United States [37]** | | | | | |
| **BMI** | | | | | |
| **3** | | | | | |
| I | −0.4 (NR) | 3.3 decrease | + | NR | NR |
| C | 0.46 (NR) | — | — | — | — |
| **6** | | | | | |
| I | −0.44 (NR) | 4.5 decrease | + + | −1.05 (SE 1.22, 90% CI −3.09 to 0.95); *P*=.39 | −0.58 (SE 0.13, 90% CI −0.84 to −0.40); *P*=.001 |
| C | 0.83 (NR) | — | — | — | — |
| **BMI z-score** | | | | | |
| **6** | | | | | |
| I | −0.18 (NR) | 28.1 decrease | + + + | −0.15 (SE 0.15, 90% CI −0.40 to 0.09); *P*=.001 | −0.12 (SE 0.03, 90% CI −0.16 to −0.07); *P*=.001 |
| C | 0.26 (NR) | — | — | — | — |
| **Jensen et al, 2019, the United States [42]** | | | | | |
| **BMI z-score** | | | | | |
| **6** | | | | | |
| I | −0.09 (NR) | 4.2 decrease | + | Not statistically significant, *P* value, NR | NR |
| C | −0.03 (NR) | — | — | — | — |
| **Love-Osborne et al, 2016, the United States [38]** | | | | | |
| **BMI z-score** | | | | | |
| **6-8** | | | | | |

| Author, year, citation, outcome, follow-up timepoint (months), and study arms | Mean difference within groups (SD) | Effect size (%) | Level of effect size[a] | Mean difference between groups | Mean difference between groups over time |
|---|---|---|---|---|---|
| I | 1.2 (NR) | 3.8 increase | + | NR | NR |
| C | 0.0 (NR) | — | — | — | — |
| **Mameli et al, 2018, Italy [39]** | | | | | |
| **BMI z-score** | | | | | |
| **3** | | | | | |
| I | NR | NR | NR | 0.00 (95% CI −0.11 to 0.12); *P*>.05 | 0.01 (95% CI −0.15 to 0.18); *P*=.87 |
| C | NR | — | — | — | — |
| **Nguyen et al, 2012, Australia [40,46]** | | | | | |
| **BMI** | | | | | |
| **2** | | | | | |
| I | NR | NR | NR | NR | NR |
| C | NR | — | — | — | — |
| **12** | | | | | |
| I | 0.6 (NR) | 1.9 increase | + | 0.1 (95% CI −1.2 to 1.3); *P*>.05 | 0.1 (95% CI −0.3 to 0.4); *P*>.05 |
| C | 0.0 (NR) | — | — | — | — |
| **24** | | | | | |
| I | 0.0 (NR) | 3.2 decrease | + | 0.1 (95% CI −1.2 to 1.3); *P*>.05 | 0.8 (95% CI 0.2 to 1.4); *P*<.05 |
| C | 1.0 (NR) | — | — | — | — |
| **BMI z score** | | | | | |
| **2** | | | | | |
| I | NR | NR | NR | NR | NR |
| C | NR | — | — | — | — |
| **12** | | | | | |
| I | −0.06 (NR) | 1.0 increase | + | −0.00 (95% CI −0.11 to 0.10); *P*>.05 | −0.09 (95% CI −0.12 to −0.06); *P*<.001 |
| C | −0.08 (NR) | — | — | — | — |
| **24** | | | | | |
| I | −0.2 (NR) | 5.4 decrease | + + | −0.01 (95% CI −0.11 to 0.10); *P*>.05 | −0.13 (95% CI −0.20 to −0.06); *P*<.05 |
| C | −0.09 (NR) | — | — | — | — |
| **Patrick et al, 2013, the United States [41]** | | | | | |
| **BMI z score** | | | | | |
| **6** | | | | | |
| W[h] | −0.1 (NR) | 4.5 decrease | + + | NR | *P*=.93 |
| WG[i] | 0.0 (NR) | — | — | — | — |
| WSMS[j] | −0.1 (NR) | — | — | — | — |
| C | 0.0 (NR) | — | — | — | — |
| **12** | | **4.5 decrease** | + + | **NR** | **—** |
| W | −0.1 (NR) | | | | |
| WG | −0.2 (NR) | | | | |

| Author, year, citation, outcome, follow-up timepoint (months), and study arms | Mean difference within groups (SD) | Effect size (%) | Level of effect size[a] | Mean difference between groups | Mean difference between groups over time |
|---|---|---|---|---|---|
| WSMS | −0.1 (NR) | | | | |
| C | 0.0 (NR) | | | | |

[a]+ denotes effect size 0%–4%; + + denotes effect size 5%–9%; + + + denotes effect size ≥10%.

[b]NR: not reported.

[c]IT: Internet intervention group.

[b]NR: not reported.

[d]sLMP: Simplified Lifestyle Modification Program intervention group.

[e]C: Control.

[f]I: Intervention.

[g]Not applicable.

[h]W: website only.

[i]WG: website + group.

[j]WSMS: website + text messages.

At 3 months, Abraham et al [35] investigated the mean difference in BMI (kg/m$^2$) between the intervention and control groups and found no effect (*P*=.04) on BMI (kg/m$^2$). At 6 months, 3 studies investigated the mean difference in BMI between the intervention and control groups and found no effect [36-38]. Two studies found significant group by time effects for mean decreases in BMI (kg/m$^2$) from baseline to 6 months [36,37]. Chen et al [36] reported a decrease in BMI of 0.58 kg/m$^2$ (SE 0.13; 90% CI −0.84 to −0.40; *P*=.001) [37] and the other study reported only a value of *P*<.001 [36]. Nguyen et al found that the intervention with adjunctive electronic contact including text messages (once per month from 2 months) compared with group-based behavioral lifestyle intervention only had no effect at 12 (BMI 0.01 kg/m$^2$; 95% CI −1.2 to 1.3; *P*>.05) [40] or 24 months (BMI 0.01 kg/m$^2$; 95% CI −1.2 to 1.3; *P*>.05) [46]. From baseline to 24 months, there was a significant group by time effect for BMI, with a mean decrease between the intervention and control groups of 0.8 kg/m$^2$ (95% CI 0.2 to 1.4; *P*<.05) [46].

At 3 months, Mameli et al [39] found comparable mean decreases in BMI z-score between the intervention and control groups (−0.03 units, 95% CI 0.14 to 0.09 vs −0.04 units, 95% CI 0.16 to 0.08, respectively), which were not significantly different (*P*>.05) and no time effect from baseline to 3 months was observed (*P*=.87). At 6 months, Chen et al [37] found a significant mean difference in BMI z-scores between groups of −0.15 units (SE 0.15; 90% CI −0.40 to 0.09; *P*=.001) and a significant decrease in BMI z-score over time between groups (−0.12 units; SE 0.03; 90% CI −0.16 to −0.07; *P*=.001). Jensen et al [42] found that BMI z-score did not differ across the intervention and control groups at 6 months (*P* value not reported). However, participants in the intervention group significantly decreased BMI z-score by 0.32 units compared with baseline values (*P*<.01), whereas the control group participants did not decrease BMI z-score over the 6 months (*P*=.63). Similar to the results for BMI, Nguyen et al [40,46] found no significant difference in BMI z-score between groups at 12 and 24 months (*P*>.05). However, from baseline to 12 and 24 months, there was a significant group by time effect for BMI z-score, with a mean decrease between the intervention and control groups of 0.09 units (95% CI −0.12 to −0.06; *P*<.001) and 0.13 units (95% CI −0.20 to −0.06; *P*<.05), respectively. Patrick et al [41] found no difference in BMI z-score from baseline, 6, and 12 months between groups (*P*=.93).

### *Lifestyle Outcomes*

There was limited overlap in secondary outcomes between studies, which limited comparisons. In total, 3 studies measured blood pressure, and no consistent significant differences were shown between the intervention and control groups. Abraham et al [35] found no decrease in systolic or diastolic blood pressure between the intervention and control groups. Chen et al [37] found a significant reduction in diastolic blood pressure of 2.66 mmHg over 6 months between the intervention and control groups (90% CI −4.02 to −1.31; *P*=.001). Whereas Nguyen et al [40] found the intervention group had a mean higher systolic blood pressure difference of 3 mmHg at 12 and 24 months only after adjusting for sex, age, and perceived athletic ability. Bagherniya et al [36] and Nguyen et al [40] found no effect of the intervention on waist circumference at 6 months or 12 and 24 months, respectively. A total of 4 studies measured physical activity using different instruments, including a 0 to 10 scale of physical activity validated in Chinese youth [35], a short question adapted from the California Health Interview Survey [37], and a 7-day physical activity recall interview [41] or physical activity questionnaire [36]. Chen et al [37] observed a significant increase of 0.40 physical activity days per week in the intervention group compared with the control group over the 6 months (90% CI 0.15 to 0.66; *P*=.01) [37] and Bagherniya et al [36] found the intervention group significantly increased daily minutes of physical activity compared with controls at 6 months (*P*<.001). Two studies assessed the quality of life using the Impact on Weight Quality of Life–Kids [42] and the pediatric quality of life inventory [41]. Jensen et al [42] found no between-group difference in the weight-related quality of life. However, both groups demonstrated significant improvements in parent-report scores (*P*s<.05) [42], and Patrick et al [41] found positive correlations between physical functioning quality of life and behavior change strategies in girls. Body fat percentage was measured in 2 studies

at 6 months, and no differences were observed [35,41]. Dietary intakes were measured in 2 studies via short self-report questions on fruit, vegetable, and sugar-sweetened beverage intake [37] or by a food frequency questionnaire [41] at 6 months, and no differences were observed. Other outcomes included waist-to-height ratio, stress, depression self-efficacy, and social support for behavior change, of which none were consistent between studies.

## Study Characteristics

All RCTs included in the review were published in English [35-41]. The median sample size was 75 participants (range 40-172 participants), with a total of 767 participants (Table 4). In all, 2 studies recruited participants from obesity clinics [35,39], 3 studies recruited from primary care [37,41,42], 2 studies recruited from schools [36,38], and the remaining study utilized several recruitment strategies [40,45] (Table 5). Participants ranged in age from 12 to 18 years, with a mean age of 14.3 years (SD 0.9). All participants had overweight (BMI on 85th-95th percentile) or obesity (BMI>95th percentile) with a mean BMI of 29.7 kg/m$^2$ (SD 1.6). One study reported the mean participant BMI percentile for age and sex at baseline, which was 91.5 (SD 4.2) [42]. On average, 53% of participants were female (range 27%-100%). Overall, 6 studies reported participant ethnicity [35-38,41,42], with 2 studies recruiting 100% of participants who were Chinese [35,37]; in 1 study, 100% of the participants were Persian [36]; in 1 study, 64% of the participants were white [42], and the samples in the remaining 2 studies were predominately Hispanic [38,41]. In all, 5 studies with a follow-up at 2 to 3 months had a median attrition rate of 6.8% (range 0%-30.4%) [35-37,39,40]. The median attrition rate increased for each subsequent follow-up: 15% at 6 to 8 months (range 0%-38.5%, n=6 studies) [35-38,41,42]; 26.9% at 12 months (range 12.3%-46.2%, n=2 studies) [40,41]; and 38.5% at 24 months (range 25.9%-41.1%, n=1 study) [40].

**Table 4.** Characteristics of included studies (n=8).

| Author, year, country | Study design | Total (n) | I[a] (n) | C (n) | Active (months)[b] | Extended (months)[b] | Follow-up(s) (months) | Attrition at follow-up(s) (%) | Dropouts compared |
|---|---|---|---|---|---|---|---|---|---|
| Abraham et al, 2015, China [35] | RCT[c] | 48 | IT[d]: 16; sLMP[e]: 16 | 16 | 3 | 3 | 3, 6 | 3 months: 0; 6 months: 0 | Not applicable |
| Bagherniya et al, 2018, Iran [36] | RCT | 172 | 87 | 85 | 7 | 0 | 3.5, 7 | 3.5 months: I: 10.3; C[f]: 2.4; 7 months: I: 16.1; C: 4.7 | Not reported |
| Chen et al, 2017, the United States [37] | RCT | 40 | 23 | 17 | 3 | 3 | 3, 6 | 3 months: I: 0; C: 0; 6 months: I: 8.7; C: 11.8 | Not reported |
| Jensen et al, 2019, the United States [42] | RCT | 47 | 29 | 18 | 6 | 0 | 6 | 6 months: I: 34; C: 33 | No difference |
| Love-Osborne et al, 2016, the United States [38] | RCT | 165 | TM[g]: 38; NTM[h]: 44 | 83 | 6-8 | 0 | 6-8 | I: 5; C: 11 | No difference |
| Mameli et al, 2018, Italy [39] | RCT | 43 | 23 | 20 | 3 | 0 | 3 | I: 30.4; C: 25 | No difference |
| Nguyen et al, 2012, Australia [40,46] | RCT | 151 | 73 | 78 | 2 | 22 | 2, 12, 24 | 2 months: I: 6.8; C: 11.5; 12 months: I: 12.3; C: 15.4; 24 months: I: 41.1; C: 35.9 | No difference |
| Patrick et al, 2013, the United States [41] | RCT | 101 | W[i]: 26; WG[j]: 26; WSMS[k]: 24 | 25 | 12 | 0 | 6, 12 | 6 months: W: 31.0; WG: 38.5; WSMS: 22.7; C: 37.5; 12 months: W: 31.0; WG: 46.2; WSMS: 22.7; C: 33.3 | No difference |

[a]I: intervention.

[b]Intervention duration.

[c]RCT: randomized controlled trial.

[d]IT: Internet intervention group.

[e]sLMP: Simplified Lifestyle Modification Program intervention group.

[f]C: control.

[g]TM: text message.

[h]NTM: no text message.

[i]W: website only.

[j]WG: website + group.

[k]WSMS: website + text messages.

**Table 5.** Characteristics of participants from included studies (n=8).

| Author, year, country | Recruitment of population | Age range, years | Age (years), mean (IQR or SD) | BMI range | BMI (kg/m$^2$), mean (IQR or SD) | Female (%) | Ethnicity, other SES[a] factors (%) |
|---|---|---|---|---|---|---|---|
| Abraham et al, 2015, China [35] | Obesity clinic | 12-18 | IT[b]: 14.9[c] (13.7-16.2); sLMP[d]: 14.1[c] (13.5-15.3); C[e]: 14.3[c] (13.5-15.8) | >95th | IT: 29.3[c] (26.7-30.9); sLMP: 31.5[c] (29.8-33.7); C: 30.1[c] (28.4-32.3) | 39.6 | Chinese: 100; Parent tertiary education: 27 |
| Bagherniya et al, 2018, Iran [36] | Government and private schools | 12-16 | I[f]: 13.5 (0.7); C: 13.4 (0.6) | ≥85th | I: 29.2 (3.9); C: 27.2 (2.9) | 100 | Persian: 100; Parent tertiary education: 100 |
| Chen et al, 2017, the United States [37] | Primary care providers | 13-18 | I: 15.0 (1.7); C: 14.8 (1.6) | ≥85th | I: 27.4 (3.3); C: 28.4 (4.4) | 42 | Chinese American: 100; Parent mean years of education: 10 |
| Jensen et al, 2019, the United States [42] | Primary care pediatric practices | 12-18 | 15.0 (1.5) | ≥85th; <95 | 91.5 (4.2) (BMI %) | 79 | Hispanic: 23; African American: 2 |
| Love-Osborne et al, 2016, the United States [38] | Public schools | 12-18 | I: 15.7 (1.5); C: 16.0 (1.5) | ≥85th | I: 31.9 (6.2); C: 31.6 (6.5) | I: 58; C: 46 | Hispanic: I: 88 and C: 89 |
| Mameli et al, 2018, Italy [39] | Obesity clinic | 10-17 | I: 12.6 (1.7[g]); C: 12.4 (2.2[g]) | ≥95th | I: 29.6 (3.3[g]); C: 28.6 (2.6[g]) | I: 31[g]; C:43[g] | Parent tertiary education: 17[g] |
| Nguyen et al, 2012, Australia [40,46] | Media, schools, health professionals and community organizations | 13-16 | I: 14.0 (0.9); C: 14.2 (1.0) | 1.0-2.5[h] | I: 30.8 (4.2); C: 30.8 (3.5) | 27 | Not reported |
| Patrick et al, 2013, the United States [41] | Pediatric primary care | 12-16 | W[i]: 14.1 (1.4); WG[j]: 14.3 (1.5); WSMS[k]: 14.3 (1.8); C: 14.5 (1.5) | >85th | W: 2.2 (0.07)[l]; WG: 2.2 (0.07)[l]; WSMS: 2.2 (0.07)[l]; C: 2.2 (0.07)[l] | 63 | Hispanic: 74 |

[a]SES: socioeconomic status.

[b]IT: Internet intervention group.

[c]Median age and BMI (IQR).

[d]sLMP: Simplified Lifestyle Modification Program intervention group.

[e]C: control.

[f]I: intervention.

[g]Completers-only data

[h]95% confidence intervals

[i]W: website only.

[j]WG: website + group.

[k]WSMS: website + text messages.

[l]BMI z-score (standard error).

## Intervention Characteristics

A total of 5 studies were two-arm RCTs [36,37,39,40,42], 2 were three-arm RCTs [35,38], and 1 a four-arm RCT (Table 5) [41]. The median active intervention period was 4.5 months (range 2-12 months) and 2 studies had an extended intervention period of 3 months [35,37], and 1 had 22 months [40]. In all, 5 studies followed up participants at 2 to 3 months [35-37,39,40], 6 studies at 6 to 7 months [35-38,41,42], 2 studies at 12 months [40,41] and 1 study at 24 months [40]. All interventions were multicomponent, and no interventions were delivered solely via text messages (Table 6). Two studies had intervention groups consisting of Web-based educational modules delivered via a website and text messages [35,41]; 2 studies had intervention groups which provided participants wearable devices and access to smartphone apps and text messages [37,39], and the remaining 4 studies had intervention groups which involved in-person individual or group sessions with health professionals and text messages [36,38,40,42]. In all, 4 studies were grounded in the social cognitive theory [35-37,40]. Overall, 2 studies were

primarily set in an obesity clinic [35,39], 2 in high schools [36,38], 3 in primary care [37,40,42], and 1 was mostly online [41]. The majority of studies encouraged peer support from parents, and all studies included interactions with research and health professionals. Control conditions consisted of usual care, with 6 of the 8 studies including control conditions with in-person educational or information sessions [35,36,39-42].

The median number of text messages sent each week was 1.5 (range 1-21 text messages; Table 7). One study specified that 3 text messages were sent each day; however, it was not specified if this was every day [42]. In this study, control participants also received a form of control text messages, and both groups demonstrated reductions in BMI z-score [42]. In all, 4 studies used one-way text messages (participants unable to reply) [36-39] and 4 studies used two-way text messages (participants able to respond) [35,40-42]. A total of 4 studies had personalized text messages [38,39,41,42], 2 studies had semipersonalized text messages [35,40], and 2 studies did not have personalized text messages [36,37]. Overall, 6 studies utilized text messages in the active intervention phase, ranging from 17% to 100% of the intervention contact points [35,36,38,39,41,42]. In all, 2 studies utilized text messages only in the extended intervention phase accounting for 100% [37] and 60% [40] of the intervention contact points. Both studies with text messages only in the extended phase had significant between-group differences in BMI or BMI z-score at the end of follow-up [37,40]. The 4 studies, which personalized the text messages, had text message content–related individual behavioral goals [35,38] or utilized data from wearables [39] or self-reported text messages [42] to provide personalized feedback on lifestyle behaviors. All 4 studies did not significantly affect BMI outcomes. The BCTs used in the text messages are presented in Multimedia Appendix 5. The number of domains ranged from 3 to 5, with all text messages in each study utilizing goals and planning and credible source. The number of BCTs ranged from 4 to 8, with all text messages in each study using goal setting for behaviors and credible source.

A total of 3 studies provided details about the development of the intervention content broadly [37] or text messages [24,35,42]. Abraham et al [35] conducted focus groups with 11 Chinese American adolescents asking open-ended questions about the use of text messages in the program, and adolescents agreed that individualized weekly text message reminders could be a good way to enhance motivation to adopt healthy behaviors. No further details about the text message development were provided. Chen et al [37] formed an advisory group consisting of two primary care physicians and four adolescents and consulted other adolescents (n=10) via a focus group. The advisory group identified the purpose and goals of the intervention and focus groups and piloted the procedures and appropriateness of the intervention content. Adolescents *liked the program* and minor changes were made. No specific details relating to text message development were provided. In a separate publication, Jensen et al [24] detailed the process of text message development. The iterative process involved the development of an evidence-based text message bank by the research team, which was then tested in a pilot study with 20 adolescents (mean age 14 years, 85% female, 77% white) who received 1 text message per day for 3 months [24]. Feedback was provided at monthly intervals and at the end of the study. Overall, participants liked the text messages, liked receiving them after school, and they could recall topics the text messages addressed. The most popular text messages were recipe ideas, followed by testimonials and messages with pictures. Most found the messages personally relevant, and messages helped them to make healthy choices and kept them focused on weight management. A greater variety of messages were suggested. The findings informed the RCT.

In total, 6 studies provided results related to the rates of replies of text messages during the intervention. Abraham et al [35] found 78.3% of participants responded to text messages about diet goals and 77.5% of participants responded to text messages about exercise goals. In a study by Nguyen et al [40], response rates were lower, with messages about healthy eating and booster session reinforcement having the highest reply rates (42% and 34%, respectively), whereas text messages about self-esteem or stress management had the lowest response rates (4%). Jensen et al [42] found over 6 months that intervention participants sent self-monitoring text messages on 47% of intervention days compared with only 22% control participants [42]. Intervention participants demonstrated significantly higher self-monitoring adherence ($P<.01$) than control participants. The 3 text messages per day were helpful for 85% of intervention participants. However, 69% of intervention participants found the text message content annoying and repetitive. Over half (54%) would prefer fewer and more personalized text messages.

Love-Osborne et al [38] sent an average of 12 personal goal-related text messages and 12 text messages to remind participants to return log sheets about their weekly weight and lifestyle behaviors. However, it was reported that only few adolescents turned in log sheets in either group. Only 1 study asked participants to rate the helpfulness of text messages, with 27 of 39 adolescents rating the messages *somewhat helpful* or *very helpful* [40]. One study, which included Fitbit and website education resource in addition to text messages, found that the 100% of participants would recommend the program to others and 91% shared the Fitbit data with their primary care provider [37]. Only 1 study detailed that text messages were sent from an automated computer system capable of tailoring the text messages using an algorithm based on participant feedback via text message [42]. No other studies provided details about the system or infrastructure used to deliver the text messages. One study reported a research assistant spent 2 hours per week sending personalized text messages [35].

**Table 6.** Characteristics of interventions (n=8).

| Author, year, country, and intervention description | Theory | Setting | Peer support | Intervention personnel interaction | Comparator description | Comparator personnel interaction |
|---|---|---|---|---|---|---|
| **Abraham et al, 2015, China [35]** | | | | | | |
| IT[a]: usual care + 12 website lessons + text messages | TTM[b]; SCT[c] | Obesity clinic; website; mobile phone | Parents | IT: usual care | Usual care: physician check-ups with obesity counselling | 3 in-person sessions for adolescents with their parent(s) |
| sLMP[d]: usual care+ 4 counselling sessions | SCT | Obesity clinic | Parents | sLMP: usual care + 4 in-person counselling sessions with nutritionist | | |
| **Bagherniya et al, 2018, Iran [36]** | | | | | | |
| I[e]: 14 sports workshops + 7 counselling sessions + up to 60 fun exercise sessions + up to 60 competitive sports sessions + up to 3 family activity sessions + text messages[f] | SCT | School; local gyms; mobile phone | Parents; teachers | I: 14 in-person sport group workshops (personnel not specified) + 7 in-person sports counselling sessions (personnel not specified) + up to 56 fun group exercise sessions with specialist in physical education | Control group: education classes, lectures, printed handbook | 3 in-person educational classes for adolescents; 2 in-person lectures for parents/teachers |
| **Chen et al, 2017, the United States [37]** | | | | | | |
| I: Fitbit Flex monitoring device + APP[g] for self-monitoring diet, physical activity + iStart Smart 8-module online education program with in-app messages + text messages (extended intervention only) | SCT | Primary care; mobile phone | Nil | I: In-person demonstrations on how to access Fitbit data and iSmart 8-module online education program | Control group: pedometer, paper food and activity diary, 8 module online program | Nil |
| **Jensen et al, 2019, the United States [42]** | | | | | | |
| I: Single in-person 50 min MI[h] session + self-monitoring and adaptive text messages | MI | Primary care | Nil | I: Single in-person MI session with a clinical physiologist doctoral student designed to elicit motivation for change, assess potential barriers to change, and reinforce weight-related behavior change talk and education on Stoplight Eating Plan | Control group: In-person MI session + self-monitoring text messages | 1 in-person MI session |
| **Love-Osborne et al, 2016, the United States [38]** | | | | | | |
| I: Health Educator visits + paper-based self-monitoring log sheet + text messages (TMs[i] subgroup of intervention group)[j] | MI | School; mobile phone | Nil | I: Up to 8 in-person visits with health educator | Not reported | Not reported |
| **Mameli et al, 2018, Italy [39]** | | | | | | |
| I: Participants provided a WB[k] + APP and asked to record the real-time food intake | Not reported | Obesity clinic; mobile phone | Parents | I: In-person training session on how to use WB and APP | Control group: Mediterranean diet advice + instruction to increase physical activity and decrease sedentary time | 1 in-person baseline information session |
| **Nguyen et al, 2012, Australia [40,46]** | | | | | | |

| Author, year, country, and intervention description | Theory | Setting | Peer support | Intervention personnel interaction | Comparator description | Comparator personnel interaction |
|---|---|---|---|---|---|---|
| I: Loozit group program of seven 75-min group sessions during active phase + seven 60-min group sessions during extended phase + ATC[l] including 32 emails or text messages+ 14 TC[m] sessions | SCT | Primary care; hospital; mobile phone | Parents | I: Active phase, 7 in-person group sessions with trained dietitians + Extended phase, 7 in-person group sessions + 32 emails or text messages + 14 TC sessions all with trained dietitians | Control group: Loozit Program only | Active phase, 7 in-person group sessions with trained dietitians + Extended phase, 7 in-person group sessions with trained dietitians |
| **Patrick et al, 2013, the United States [41]** | | | | | | |
| WSMS[n]: W[o] + text messages | BDM[p]; TTM | Website; mobile phone | Nil | WSMS: communication with health counsellor via text message | Control group: printed educational material + encouraged to attend three 60-min group nutrition sessions at local hospital + monthly mailed tip sheets | Up to 3 in-person group nutrition sessions |
| WG[q]: W + 12 90-min group sessions + 24 health coaching calls | BDM; TTM | Website | Parents | WG: In-person group sessions with health counsellor | Not reported | Not reported |
| W: wkly check-in emails + monthly mailed tip sheets + access to program website + website tutorials | BDM; TTM | Website | Nil | W: communication with health counsellor via email | Not reported | Not reported |

[a]IT: Internet intervention group.

[b]TTM: transtheoretical model.

[c]SCT: social cognitive theory.

[d]sLMP: Simplified Lifestyle Modification Program intervention group.

[e]I: intervention.

[f]In addition, parent text messages and newsletters and increased sports equipment in schools.

[g]APP: smartphone app.

[h]MI: motivational interviewing.

[i]TMs: text messages group.

[j]All participants in the intervention group (TMs and NTMs) were offered text messages during the second school semester.

[k]WB: wristband.

[l]ATC: additional therapeutic contact.

[m]TC: telephone coaching.

[n]WSMS: website + text messages.

[o]W: website only.

[p]BDM: behavioral determinants model.

[q]WG: website + group.

[r]Separate sessions for adolescents and parents or caregivers.

**Table 7.** Text message details of included studies (n=8).

| Author, year, country | Duration, n | Direction | Personalized | Content | Intervention (%)[a] |
|---|---|---|---|---|---|
| Abraham et al, 2015, China [35] | 1/week | Two-way | Semi | Individual diet and exercise goals | Active: 46; Extended: 100 |
| Bagherniya et al, 2018, Iran [36] | 1/week | One-way | No | Main goals of program, strategies to overcome barriers | Active: 17 |
| Chen et al, 2017, the United States [37] | 2/week | One-way | No | Reinforced adoption and maintenance of healthy lifestyles and weight management practices | Active: 0; Extended: 100 |
| Jensen et al, 2019, the United States [42] | 3/day | Two-way | Yes | Self-monitoring text messages: participants sent text messages reporting 4 behaviors; Adaptive text messages: Evidenced-based intervention content delivered in a gain-frame format (indicating what might be gained from adopting healthier behaviors) | Active: 100 |
| Love-Osborne et al, 2016, the United States [38] | 2/week | One-way | Yes | One individualized goal-related text message and one reminder to return self-monitoring log sheet | Active: 93 |
| Mameli et al, 2018, Italy [39] | 1/week | One-way | Yes | Using previous 7-day WB[b] and app data, feedback about dietary compliance and quality, energy gap, sedentary time, physical activity and suggestions on how to reach each of 5 goals | Active: 100 |
| Nguyen et al, 2012, Australia [40,46] | 1/month | Two-way | Semi | Reinforce key healthy lifestyle principles covered during the active phase and extended phase group sessions | Active: 0; Extended: 60 |
| Patrick et al, 2013, the United States [41] | 3/week | Two-way | Yes | Related to weekly challenges and intervention goals | Active: 75 |

[a]Percentage of intervention delivered by text message was determined by (number of text message contact points ÷ total number of contact points)×100. Each in-person session, website session, or text message is counted as one contact point.

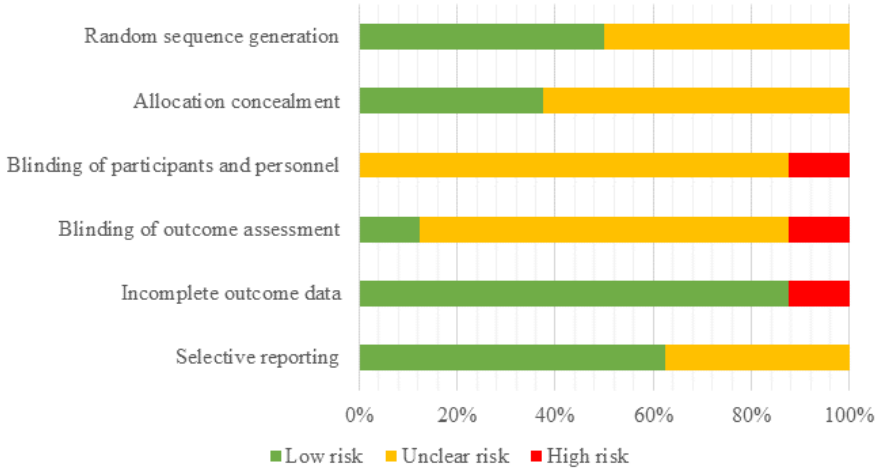
[b]WB: wristband.

## Risk of Bias Assessment

### *Cochrane Risk of Bias Assessment*

Overall, the majority of studies were classified as having a low or unclear risk of bias across all main sources of bias (Figure 2 and Multimedia Appendix 6). In all, 5 studies described a random component in the sequence generation process, permitting a judgment of low risk of bias for random sequence generation [35,37,39,40,42]; 5 studies provided insufficient information to determine if the allocation was adequately concealed [36-39,41]; and 7 studies did not provide adequate information to determine if there was blinding of participants and personnel to the intervention [35-38,40-42]. One study stated that there was no blinding and was classified as high risk [39]. One study has blinded outcome assessment [36], 1 study had no blinding of outcome assessment [39], and the remaining studies were unclear [35,37,38,40-42]. A total of 7 studies adequately addressed incomplete data permitting a judgment of low risk of bias [35,37-42]. Overall, 4 studies were rated low risk of bias for selected reporting as they provided a study protocol or trial registration which prespecified all outcome [35-37,40].

**Figure 2.** Risk of bias assessment summary.




### *Grading of Recommendations Assessment, Development, and Evaluation Overall Assessment*

The search attempted to identify all interventions for weight loss or management in adolescents delivered via text messages to answer the research question. In the body of studies identified (n=8), there were limitations in study design, variant populations, complex interventions, and potential publication bias resulting in an overall low-quality rating (Table 8).

**Table 8.** Overall assessment of quality in 8 studies (767 participants in total) of weight loss or weight management interventions in adolescents delivered via text message using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system.

| Category | Rating with reason |
|---|---|
| Study limitations | −2 quality levels due to very serious limitations |
| Consistency | No subtraction in levels |
| Directness | −1 quality level, as the interventions are indirect |
| Precision | No subtraction of levels due to good precision |
| Publication bias | −1 quality level, as publication bias cannot be ruled out |
| Overall quality | Low: our confidence in the effect estimate is limited |

### *Study Limitations*

All 8 included studies were RCTs. However, within each of the primary sources of bias, there was insufficient information to permit judgments. In all, 3 studies did not provide sufficient detail on the random sequence generation [36,38,41], and 5 studies did not provide adequate detail on the allocation concealment to permit judgment based on selection bias [36-39,41]. A total of 7 studies did not provide sufficient detail of blinding methods for participants and personnel [35-38,40-42], and 1 study did not blind participants and personnel [39]. The blinding of the outcome assessment was unclearly described in 6 studies [35,37,38,40-42], and 1 study did not blind participants or personnel to the outcome assessment [39]. Attrition bias was rated as low risk in 7 studies, and the median attrition rate was below 20% at 3 and 6 months [35,37-41]. However, attrition rates increased above 20% for the 2 studies with 12- and 24-month follow-ups [40,41]. Reporting bias was rated as low risk in 4 studies [35-37,40] and unclear in 4 studies due to insufficient information reported in the publication [38,39,41,42], further limiting the quality of the body of evidence.

### *Consistency*

Overall, 4 of the 5 studies found the interventions with text messages did not significantly decrease BMI in intervention participants compared with control participants. Of 5 studies, 3 found the interventions with text messages did not significantly decrease BMI z-score in intervention participants compared with control participants. This result suggests a trend toward no effect. However, no studies reported any subgroup analyses on the effect of text messages only. Therefore, it cannot be determined if text messages were responsible for the intervention effect on BMI or BMI z-score.

### *Directness*

Comparisons between the control and intervention arms were direct for the included interventions; variations in the study design, populations, and interventions resulted in the overall body of evidence were indirect. The population of all interventions were adolescents, 10 to 19 years. However, 3

studies recruited all adolescents who were Chinese [35], Chinese American [37], or Persian [36], 2 studies recruited adolescents who were predominately Hispanic [38,41], 1 study recruited adolescents who were predominately white [42], and 2 studies did not report ethnicity of adolescent participants [39,40]. Only 4 studies reported the parents' education level (range 17%-100% with tertiary education) [35-37,39]. The population is not representative of the broader adolescent population.

This review allowed for the inclusion of studies that used text messages as a component of a multicomponent intervention. Consequently, all studies used text messages in combination with other intervention components. In addition, 2 studies only used text messages during the extended intervention phase [37,40], and no studies evaluated the effectiveness of only the text message intervention delivery with respect to BMI. All studies measured BMI according to standardized procedures. However, the overall evidence is an indirect representation of the effect of weight management or weight loss interventions delivered via text messages on BMI.

### *Precision*

A total of 6 studies reported power calculations and 3 studies were pilot or feasibility studies. All power calculations were based on the primary outcome of BMI (measured as BMI kg/m$^2$ or BMI z-score). Sample size varied from 40 to 172 participants, with a total of 767 participants, which is considered a sufficient body of evidence for addressing the research question.

### *Publication Bias*

A comprehensive search was conducted to ensure that all the available literature was retrieved. This search included major electronic databases, hand searching of reference lists, clinical trial databases and the gray literature, and contacting authors for additional information. Also, as text messages often formed part of larger multicomponent interventions, a conservative approach for full-text screening was adopted to ensure all possible studies were included. However, we may have missed unpublished interventions with insignificant results or negative findings. Therefore, publication bias cannot be ruled out.

## Discussion

### Principal Findings

In this review, we assessed 8 studies that investigated the effectiveness of interventions delivered by text messages for weight management in adolescent populations with overweight or obesity. Of the 8 studies, 7 demonstrated reductions in BMI or BMI z-score in the intervention group compared with the control group at the end of the final follow-up. The effect was only statistically significant in 1 study at 6 months [37]. Significant time by group effects were observed in 3 studies [36,37,40]. The intervention characteristics were heterogenous and multicomponent. However, in the 8 studies, text messages accounted for over 50% of intervention contact points during the active phase, and for the 3 studies with an extended intervention phase, text messages accounted for over 85% of the intervention contact points. Despite text messages accounting for large proportions of the intervention, it cannot be determined if the effects are attributable to the text messages, as no studies utilized a factorial study design and limited process evaluation data to elucidate the characteristics and effects of only text message intervention delivery on BMI outcomes. Also, the overall quality of the body of evidence was rated as low, and this restricts the conclusions that can be drawn. Taking these findings together, this review found that there is limited high-quality evidence for the effectiveness of text messages for weight management in adolescents with overweight or obesity; we, therefore, propose suggestions for improved research design.

### Comparison With Prior Work

There were only 8 RCT studies included in this review, which utilized text messages for weight management intervention delivery in adolescent populations. The small evidence base may indicate that digital research strategies in this area is in its infancy. However, it also well-recognized that investments into research for adolescent obesity prevention have not kept pace with population growth [48]. Many of the successes in childhood obesity prevention over recent times have been the result of targeted investment in interventions benefiting younger children [49,50]. Investing in obesity prevention and management interventions into adolescence is critical to consolidating these successes. A 2019 Cochrane review of 153 RCTs testing the effectiveness of a range of interventions that include diet or physical activity components, or both, designed to prevent obesity in children found only 20% (31/153) of studies targeted adolescents aged 13 to 18 years [10]. Over 50% (85/153) of interventions were targeted at children aged 6 to 12 years. The review concluded that strategies for changing diet or physical activity levels, or both, of children to help prevent them from becoming overweight or obese are effective in making modest reductions in BMI z-score in children aged 6 to 12 years. Similar to the finding of the current systematic review, the Cochrane review found limited evidence for adolescents aged 13 to 18 years, and the diet and physical activity strategies given to them in the studies did not reduce their BMI z-score [10]. Interventions to prevent obesity, including combinations of both physical activity and diet interventions, have so far remained mainly school based and with limited evidence of effective digital intervention strategies. A 2017 systematic review, which focused on digital interventions for improving the diet and physical activity behaviors of adolescents, identified only 4 text message interventions [26]. The 4 interventions were heterogenous and were ineffective at improving physical activity and dietary behaviors [26].

The interventions included in this review were also heterogenous, and in the most part, text messages formed a part of larger, complex, and multicomponent interventions. A systematic review by Moores et al [51], which included all study types of community-based interventions for the treatment of overweight and obesity in adolescents, found that digital technology was utilized in less than half of the 21 included interventions for adolescents, and text messages were only utilized as a part of larger interventions in 4 studies in total. The authors concluded that the inclusion of digital technology did not improve program effectiveness. Similar to the studies included in this review, the interventions were evaluated in traditional RCTs as an aggregation of components. As the

traditional RCT evaluates the intervention only as a whole, this evaluation did not enable the isolation of the individual effects of the text messages. Greater consideration in intervention design is required to understand individual user preferences and their engagement with the different intervention components, particularly newer digital technologies. In future studies, this may be achieved through the creation and evaluation of an adaptive intervention using factorial research designs, such as the sequential multiple assessment randomized trial study designs [52]. Post hoc process evaluations were conducted in 5 of the 8 studies included in this review, which allowed some understanding of individual user preferences and how adolescents engaged with and used the text messages. The 3 studies which involved adolescents in the text message development found high adherence rates measured by text message responses and high overall approval ratings [35,37,42].

Adolescents are digital frontrunners and their lack of input into digital technologies to manage their health and well-being is likely to result in ineffective levels of engagement and, subsequently, ineffective interventions [48]. Only 3 of the 8 studies included in this review engaged adolescents in the text message development [35,37,42]. The most comprehensive text message development with adolescents was conducted by Jensen et al [42], who before the start of the intervention, tested the text messages in a 3-month feasibility study and surveyed adolescents about the text message content, timings, and interactivity at monthly intervals and semistructured interviews at the end of the study [24]. This participatory development process allowed the researchers to develop a text message bank that was engaging for participants in the intervention study, with 85% of participants finding the text messages helpful. Similarly, another feasibility study utilized extensive participatory methods with adolescents to develop 300 healthy lifestyle messages and a delivery protocol [53]. The participatory partnership with adolescents allowed researchers to understand adolescents' preferred text message tone and appreciate adolescents' sensitivity to the language. For example, messages that used an authoritarian tone were strongly disliked by adolescents. This finding was also highlighted in a process evaluation of a 6-month maintenance text message intervention with adolescents with overweight or obesity [54]. The text message program, which followed an 8-week in-person intervention, was developed without adolescent consultation, and the adolescents described a sense of shame when receiving the text messages and parents perceived the text messages as impersonal. Thus, co-design may increase the likelihood of acceptable text messages and, subsequently, may result in effective engagement in future interventions for adolescents.

Other characteristics related to text message intervention engagement include dose and interactivity. Of the 8 interventions in the current review, 4 had a low dose of text messages of once per month [40] or once per week [35,36,39]. Moreover, only 4 of the included studies utilized two-way text message communication, which allowed participants to have varying levels of interaction with research staff or health professionals [35,40-42]. Jensen et al [42] requested both groups self-report behaviors to the research team via text message [42]. Participants in the intervention group, who also received adaptive text messages (evidenced-based intervention content delivered in a gain-frame format indicating what might be gained from adopting healthier behaviors), demonstrated significantly greater self-reporting adherence compared with the control group ($P<.01$). Comparably, a text messaging program on adolescent reproductive health, which compared one-way text messages with reproductive health information vs interactive text messages with reproductive health quizzes, found interactive text messages increased reproductive knowledge by 24% vs 11% in the one-way text message group [55]. The interactive text message group was instead sent one multiple-choice quiz question each week. When the participant responded, they immediately received a confirmatory text message informing them whether they answered correctly along with the correct answer and additional information, which corresponded to the information provided in the unidirectional text message. Moreover, a momentary ecological assessment using text messages to assess the adolescent's health information needs found adolescents are willing to use text messaging to report their health information in real time [56]. The research team sent 3 text messages per week asking adolescents if they had any questions about their health and where did they look for the answer. Adolescents responded to 90% of the 3 text messages sent each week by the research team. Adolescents' most frequently reported questions were about diet and exercise, and adolescents had heightened awareness and health information needs regarding issues related to obesity. Importantly, this sample of adolescents was from a lower socioeconomic status, and text messages showed promise for engaging at-risk and hard-to-reach populations. Future interventions may consider interactive two-way text messages as an intervention delivery strategy for adolescent populations given the demonstration of acceptability in formative research. Unlike other digital technologies, such as online websites and smartphone apps, text messages are inexpensive to develop and to send and receive, and they do not require an internet connection.

To determine the effectiveness of text message interventions for adolescent weight management, consistent outcome measures of body weight are required. Moreover, consistency in secondary outcome measures is essential for behaviors that underlie weight management, such as diet and physical activity behaviors, which have independent effects of their own on health [57]. In the 8 included studies, the primary outcome of interest was weight change measured either using BMI z-score or BMI. BMI z-score is a widely accepted way of assessing a child's weight status using measures of relative weight adjusted for child age and sex in their country. Of the 5 studies which reported BMI z-score, 2 studies provided insufficient details (groups means at baseline and follow-up time points) to determine an effect size [39,42] and 3 studies only reported BMI [35,36,38]. In growing adolescents, BMI varies with age and sex. As a result, for BMI to be meaningful in adolescents, it must be compared with a reference-standard that accounts for child age and sex [58]. Studies which report outcomes in BMI usually have small effect sizes compared with studies reporting BMI z-scores [51,59]. For example, Chen et al [37] had a BMI effect size of 4.5% at 6 months and BMI z-score effect size of 28.1% at 6

months [37]. This difference in outcome measures limits a quantitative comparison in effect between studies.

### Strengths and Limitations

The strengths of this systematic review include the use of a comprehensive search strategy, developed in conjunction with an experienced librarian. A thorough full-text search was undertaken as text messages often form part of larger multicomponent interventions and all publications in any language were considered. Moreover, a comprehensive quality assessment was undertaken using the Cochrane Risk of Bias tool and GRADE tool. Emails were sent to corresponding authors to obtain missing details and for clarification of text message implementation and outcome measures. Hence, more information was obtained than was published. However, several methodological limitations in this review restrict the conclusions that can be drawn. No studies were identified, which were text message–only interventions. This limited the conclusions that were drawn from the review. The risk of bias and methodological quality of the studies in the review was low and reflected the exploratory nature of text message–delivered interventions as research in this area is in its infancy. Only RCTs were included in the current review, and small uncontrolled studies were excluded, which may bias the results. Moreover, due to heterogenous intervention designs and missing outcome data, the sample size was too small for a meta-analysis of these outcomes to be conducted at each follow-up time point.

### Conclusions

This review demonstrates that there is limited evidence to suggest text messages are an effective tool to deliver interventions for weight management in adolescents with overweight or obesity. Interventions delivered by text message for adolescent weight management were heterogeneous, and the small number of studies indicates research in this area is in its infancy. Although very little high-quality research has been done in the adolescent population, there is high-quality research in adult populations that suggests text messages are an effective tool for the delivery of chronic disease management programs. Further research is required to elucidate the effectiveness and potential impact of only text message interventions on weight and weight-related behaviors in adolescents. Specifically, high-quality studies that evaluate text messages as the only modality of intervention delivery are required.

## Acknowledgments

The authors would like to thank the University of Sydney librarian, Edward Luca for his assistance with setting up the database search strategy. The authors disclose receipt of the following financial support for this research: SP is cofunded by a National Health and Medical Research Council/National Heart Foundation Early Career Fellowship (APP1157438); AS is funded by the University of Sydney's Research Training Program Scholarship and the Westmead Applied Research Centre's Westmead Applied Research Centre Supplementary Postgraduate Research Scholarship in Breast Cancer; KH is funded by a National Heart Foundation Postdoctoral Fellowship (102138); and JR is funded by a National Health and Medical Research Council Career Development Fellowship (APP1143538).

## Authors' Contributions

SP, RR, AS, KH, and JF developed the research question. SP drafted the manuscript, and RR assisted with screening, extraction, data tabulation, quality assessment, and interpretation of findings. KH provided statistical advice regarding quantitative synthesis. All authors have read and contributed to the final manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analysis checklist.
[DOC File , 64 KB - mhealth_v8i5e15849_app1.doc ]

Multimedia Appendix 2
Full electronic search strategies for each database.
[DOCX File , 42 KB - mhealth_v8i5e15849_app2.docx ]

Multimedia Appendix 3
Screenshots of full electronic search strategies for each.
[DOCX File , 3875 KB - mhealth_v8i5e15849_app3.docx ]

Multimedia Appendix 4
Full text excluded and the reason for exclusion (n=201).
[DOCX File , 57 KB - mhealth_v8i5e15849_app4.docx ]

Multimedia Appendix 5
Behaviour change techniques, text message development and process measures in individual included studies (n=8).
[DOCX File , 29 KB - mhealth_v8i5e15849_app5.docx ]

Multimedia Appendix 6
The Cochrane Collaboration's tool for assessing risk of bias in individual included studies (n=8).
[DOCX File , 24 KB - mhealth_v8i5e15849_app6.docx ]

## References


1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014 Aug 30;384(9945):766-781 [FREE Full text] [doi: 10.1016/S0140-6736(14)60460-8] [Medline: 24880830]
2. Zheng Y, Manson JE, Yuan C, Liang MH, Grodstein F, Stampfer MJ, et al. Associations of weight gain from early to middle adulthood with major health outcomes later in life. J Am Med Assoc 2017 Jul 18;318(3):255-269 [FREE Full text] [doi: 10.1001/jama.2017.7092] [Medline: 28719691]
3. Dahm CC, Chomistek AK, Jakobsen MU, Mukamal KJ, Eliassen AH, Sesso HD, et al. Adolescent diet quality and cardiovascular disease risk factors and incident cardiovascular disease in middle-aged women. J Am Heart Assoc 2016 Dec 20;5(12):pii: e003583 [FREE Full text] [doi: 10.1161/JAHA.116.003583] [Medline: 27998915]
4. Barker AR, Gracia-Marco L, Ruiz JR, Castillo MJ, Aparicio-Ugarriza R, González-Gross M, et al. Physical activity, sedentary time, TV viewing, physical fitness and cardiovascular disease risk in adolescents: The HELENA study. Int J Cardiol 2018 Mar 1;254:303-309. [doi: 10.1016/j.ijcard.2017.11.080] [Medline: 29221862]
5. World Health Organization.: World Health Organization Health for the World's Adolescents: A Second Chance in the Second Decade URL: http://apps.who.int/adolescent/second-decade/ [accessed 2017-12-12]
6. Patton GC, Viner R. Pubertal transitions in health. Lancet 2007 Mar 31;369(9567):1130-1139. [doi: 10.1016/S0140-6736(07)60366-3] [Medline: 17398312]
7. Lascar N, Brown J, Pattison H, Barnett AH, Bailey CJ, Bellary S. Type 2 diabetes in adolescents and young adults. Lancet Diabetes Endocrinol 2018 Jan;6(1):69-80. [doi: 10.1016/S2213-8587(17)30186-9] [Medline: 28847479]
8. Pont SJ, Puhl R, Cook SR, Slusser W, Section on Obesity, Obesity Society. Stigma experienced by children and adolescents with obesity. Pediatrics 2017 Dec;140(6):pii: e20173034 [FREE Full text] [doi: 10.1542/peds.2017-3034] [Medline: 29158228]
9. Schwimmer JB, Burwinkle TM, Varni JW. Health-related quality of life of severely obese children and adolescents. J Am Med Assoc 2003 Apr 9;289(14):1813-1819. [doi: 10.1001/jama.289.14.1813] [Medline: 12684360]
10. Brown T, Moore TH, Hooper L, Gao Y, Zayegh A, Ijaz S, et al. Interventions for preventing obesity in children. Cochrane Database Syst Rev 2019 Jul 23;7:CD001871. [doi: 10.1002/14651858.CD001871.pub4] [Medline: 31332776]
11. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]
12. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile health interventions for improving health outcomes in youth: a meta-analysis. JAMA Pediatr 2017 May 1;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
13. Lenhart A. Pew Research Centre. 2015 Apr 9. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/# [accessed 2017-12-06]
14. Roy Morgan Research. 2016 Aug 22. 9 in 10 Aussie Teens Now Have a Mobile URL: http://bit.ly/2bbuoAX [accessed 2017-09-20]
15. Byun YH, Ha M, Kwon HJ, Choi KH, Burm E, Choi Y, et al. Epidemiological characteristics of mobile phone ownership and use in korean children and adolescents. Environ Health Toxicol 2013;28:e2013018 [FREE Full text] [doi: 10.5620/eht.2013.28.e2013018] [Medline: 24498595]
16. Park E, Kwon M. Health-related internet use by children and adolescents: systematic review. J Med Internet Res 2018 Apr 3;20(4):e120 [FREE Full text] [doi: 10.2196/jmir.7731] [Medline: 29615385]
17. Chow CK, Redfern J, Hillis GS, Thakkar J, Santo K, Hackett ML, et al. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. J Am Med Assoc 2015;314(12):1255-1263. [doi: 10.1001/jama.2015.10945] [Medline: 26393848]
18. Bobrow K, Farmer AJ, Springer D, Shanyinde M, Yu L, Brennan T, et al. Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-Text Adherence Support [StAR]): a single-blind, randomized trial. Circulation 2016 Feb 9;133(6):592-600 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.017530] [Medline: 26769742]

19. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res 2007 Apr 27;9(2):e7 [FREE Full text] [doi: 10.2196/jmir.9.2.e7] [Medline: 17478409]
20. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
21. Siopis G, Chey T, Allman-Farinelli M. A systematic review and meta-analysis of interventions for weight management using text messaging. J Hum Nutr Diet 2015 Feb;28(Suppl 2):1-15. [doi: 10.1111/jhn.12207] [Medline: 24480032]
22. Buis LR, Hirzel L, Turske SA, des Jardins TR, Yarandi H, Bondurant P. Use of a text message program to raise type 2 diabetes risk awareness and promote health behavior change (part I): assessment of participant reach and adoption. J Med Internet Res 2013 Dec 19;15(12):e281 [FREE Full text] [doi: 10.2196/jmir.2928] [Medline: 24356329]
23. Woolford SJ, Barr KL, Derry HA, Jepson CM, Clark SJ, Strecher VJ, et al. OMG do not say LOL: obese adolescents' perspectives on the content of text messages to enhance weight loss efforts. Obesity (Silver Spring) 2011 Dec;19(12):2382-2387 [FREE Full text] [doi: 10.1038/oby.2011.266] [Medline: 21869762]
24. Woolford SJ, Clark SJ, Strecher VJ, Resnicow K. Tailored mobile phone text messages as an adjunct to obesity treatment for adolescents. J Telemed Telecare 2010;16(8):458-461 [FREE Full text] [doi: 10.1258/jtt.2010.100207] [Medline: 20959393]
25. Keating SR, McCurry MK. Systematic review of text messaging as an intervention for adolescent obesity. J Am Assoc Nurse Pract 2015 Dec;27(12):714-720. [doi: 10.1002/2327-6924.12264] [Medline: 25914399]
26. Rose T, Barker M, Jacob CM, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
27. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec;10(6):403-409 [FREE Full text] [doi: 10.1111/ijpo.12002] [Medline: 25641770]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
29. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
30. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
31. Kolsgaard ML, Joner G, Brunborg C, Anderssen SA, Tonstad S, Andersen LF. Reduction in BMI z-score and improvement in cardiometabolic risk factors in obese children and adolescents. The Oslo Adiposity Intervention Study - a hospital/public health nurse combined treatment. BMC Pediatr 2011 May 27;11:47 [FREE Full text] [doi: 10.1186/1471-2431-11-47] [Medline: 21619652]
32. Ford AL, Hunt LP, Cooper A, Shield JP. What reduction in BMI SDS is required in obese adolescents to improve body composition and cardiometabolic health? Arch Dis Child 2010 Apr;95(4):256-261. [doi: 10.1136/adc.2009.165340] [Medline: 19966092]
33. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. Br Med J 1997 Dec 6;315(7121):1533-1537 [FREE Full text] [doi: 10.1136/bmj.315.7121.1533] [Medline: 9432252]
34. Schünemann H, Brożek J, Guyatt G, Oxman A. RADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013;guidelinedevelopment.org/handbook. 2013. URL: https://gdt.gradepro.org/app/handbook/handbook.html [accessed 2020-03-30]
35. Abraham AA, Chow W, So H, Yip BH, Li AM, Kumta SM, et al. Lifestyle intervention using an internet-based curriculum with cell phone reminders for obese Chinese teens: a randomized controlled study. PLoS One 2015;10(5):e0125673 [FREE Full text] [doi: 10.1371/journal.pone.0125673] [Medline: 25946465]
36. Bagherniya M, Darani FM, Sharma M, Maracy MR, Birgani RA, Ranjbar G, et al. Assessment of the efficacy of physical activity level and lifestyle behavior interventions applying social cognitive theory for overweight and obese girl adolescents. J Res Health Sci 2018 Apr 7;18(2):e00409 [FREE Full text] [Medline: 29784890]
37. Chen JL, Guedes CM, Cooper BA, Lung AE. Short-term efficacy of an innovative mobile phone technology-based intervention for weight management for overweight and obese adolescents: pilot study. Interact J Med Res 2017 Aug 2;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
38. Love-Osborne K, Fortune R, Sheeder J, Federico S, Haemer MA. School-based health center-based treatment for obese adolescents: feasibility and body mass index effects. Child Obes 2014 Oct;10(5):424-431. [doi: 10.1089/chi.2013.0165] [Medline: 25259781]

39. Mameli C, Brunetti D, Colombo V, Bedogni G, Schneider L, Penagini F, et al. Combined use of a wristband and a smartphone to reduce body weight in obese children: randomized controlled trial. Pediatr Obes 2018 Feb;13(2):81-87. [doi: 10.1111/ijpo.12201] [Medline: 27900849]
40. Nguyen B, Shrewsbury VA, O'Connor J, Steinbeck KS, Lee A, Hill AJ, et al. Twelve-month outcomes of the loozit randomized controlled trial: a community-based healthy lifestyle program for overweight and obese adolescents. Arch Pediatr Adolesc Med 2012 Feb;166(2):170-177. [doi: 10.1001/archpediatrics.2011.841] [Medline: 22312175]
41. Patrick K, Norman GJ, Davila EP, Calfas KJ, Raab F, Gottschalk M, et al. Outcomes of a 12-month technology-based intervention to promote weight loss in adolescents at risk for type 2 diabetes. J Diabetes Sci Technol 2013 May 1;7(3):759-770 [FREE Full text] [doi: 10.1177/193229681300700322] [Medline: 23759410]
42. Jensen CD, Duraccio KM, Barnett KA, Fortuna C, Woolford SJ, Giraud-Carrier CG. Feasibility, acceptability, and preliminary effectiveness of an adaptive text messaging intervention for adolescent weight control in primary care. Clin Pract Pediatr Psychol 2019;7(1):57-67. [doi: 10.1037/cpp0000268]
43. Chen JL, Guedes CM, Lung AE. Smartphone-based healthy weight management intervention for Chinese American adolescents: short-term efficacy and factors associated with decreased weight. J Adolesc Health 2019 Apr;64(4):443-449. [doi: 10.1016/j.jadohealth.2018.08.022] [Medline: 30409751]
44. Kornman KP, Shrewsbury VA, Chou AC, Nguyen B, Lee A, O'Connor J, et al. Electronic therapeutic contact for adolescent weight management: the Loozit study. Telemed J E Health 2010;16(6):678-685. [doi: 10.1089/tmj.2009.0180] [Medline: 20575613]
45. Nguyen B, McGregor K, O'Connor J, Shrewsbury V, Lee A, Steinbeck K, et al. Recruitment challenges and recommendations for adolescent obesity trials. J Paediatr Child Health 2012 Jan;48(1):38-43. [doi: 10.1111/j.1440-1754.2011.02183.x] [Medline: 22250828]
46. Nguyen B, Shrewsbury VA, O'Connor J, Steinbeck KS, Hill AJ, Shah S, et al. Two-year outcomes of an adjunctive telephone coaching and electronic contact intervention for adolescent weight-loss maintenance: the Loozit randomized controlled trial. Int J Obes (Lond) 2013 Mar;37(3):468-472. [doi: 10.1038/ijo.2012.74] [Medline: 22584456]
47. Bagherniya M, Sharma M, Mostafavi F, Keshavarz SA. Application of social cognitive theory in predicting childhood obesity prevention behaviors in overweight and obese Iranian adolescents. Int Q Community Health Educ 2015;35(2):133-147. [doi: 10.1177/0272684X15569487] [Medline: 25856805]
48. World Health Organization. Engaging Young People for Health and Sustainable Development: Strategic Opportunities for the World Health Organization and Partners. Geneva: World Health Organization; 2018.
49. Mihrshahi S, Gow ML, Baur LA. Contemporary approaches to the prevention and management of paediatric obesity: an Australian focus. Med J Aust 2018;209(6):267-274. [doi: 10.5694/mja18.00140]
50. Redfern J, Enright G, Hyun K, Raadsma S, Allman-Farinelli M, Innes-Hughes C, et al. Effectiveness of a behavioural incentive scheme linked to goal achievement in overweight children: a cluster randomized controlled trial. J Obes Diabetes 2019;3:1-9. [doi: 10.33805/2638-812x.109]
51. Moores CJ, Bell LK, Miller J, Damarell RA, Matwiejczyk L, Miller MD. A systematic review of community-based interventions for the treatment of adolescents with overweight and obesity. Obes Rev 2018 May;19(5):698-715. [doi: 10.1111/obr.12660] [Medline: 29345042]
52. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
53. Hingle M, Nichter M, Medeiros M, Grace S. Texting for health: the use of participatory methods to develop healthy lifestyle messages for teens. J Nutr Educ Behav 2013;45(1):12-19. [doi: 10.1016/j.jneb.2012.05.001] [Medline: 23103255]
54. Smith KL, Kerr DA, Fenner AA, Straker LM. Adolescents just do not know what they want: a qualitative study to describe obese adolescents' experiences of text messaging to support behavior change maintenance post intervention. J Med Internet Res 2014 Apr 8;16(4):e103 [FREE Full text] [doi: 10.2196/jmir.3113] [Medline: 24713407]
55. Rokicki S, Cohen J, Salomon JA, Fink G. Impact of a text-messaging program on adolescent reproductive health: a cluster-randomized trial in Ghana. Am J Public Health 2017 Feb;107(2):298-305. [doi: 10.2105/AJPH.2016.303562] [Medline: 27997236]
56. Schnall R, Okoniewski A, Tiase V, Low A, Rodriguez M, Kaplan S. Using text messaging to assess adolescents' health information needs: an ecological momentary assessment. J Med Internet Res 2013 Mar 6;15(3):e54 [FREE Full text] [doi: 10.2196/jmir.2395] [Medline: 23467200]
57. GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1345-1422 [FREE Full text] [doi: 10.1016/S0140-6736(17)32366-8] [Medline: 28919119]
58. Must A, Anderson SE. Body mass index in children and adolescents: considerations for population-based applications. Int J Obes (Lond) 2006 Apr;30(4):590-594. [doi: 10.1038/sj.ijo.0803300] [Medline: 16570087]

59. Lloyd-Richardson EE, Jelalian E, Sato AF, Hart CN, Mehlenbeck R, Wing RR. Two-year follow-up of an adolescent behavioral weight control intervention. Pediatrics 2012 Aug;130(2):e281-e288 [FREE Full text] [doi: 10.1542/peds.2011-3283] [Medline: 22753560]

## Abbreviations

**BCTs:** behavioral techniques
**CVD:** cardiovascular disease
**GRADE:** Grading of Recommendations Assessment, Development, and Evaluation
**mHealth:** mobile health
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis
**RCTs:** randomized controlled trials

*Edited by G Eysenbach; submitted 12.08.19; peer-reviewed by L Lytle, J Job, S Smith; comments to author 08.12.19; revised version received 19.12.19; accepted 10.02.20; published 26.05.20.*

*Please cite as:*
*Partridge SR, Raeside R, Singleton A, Hyun K, Redfern J*
*Effectiveness of Text Message Interventions for Weight Management in Adolescents: Systematic Review*
*JMIR Mhealth Uhealth 2020;8(5):e15849*
*URL: http://mhealth.jmir.org/2020/5/e15849/*
*doi:10.2196/15849*
*PMID:32348264*

Original Paper

# An App for Detecting Bullying of Nurses Using Convolutional Neural Networks and Web-Based Computerized Adaptive Testing: Development and Usability Study

Shu-Ching Ma[1,2*], PhD; Willy Chou[3,4*], MD; Tsair-Wei Chien[5*], MBA; Julie Chi Chow[6,7], MD; Yu-Tsen Yeh[8], BA; Po-Hsin Chou[9,10], MD; Huan-Fang Lee[11*], PhD

[1]Department of Nursing, Chi Mei Medical Center, Tainan, Taiwan
[2]College of Humanities and Social Science, Southern Taiwan University of Science and Technology, Tainan, Taiwan
[3]Department of Physical Medicine and Rehabilitation, Chi Mei Medical Center, Tainan, Taiwan
[4]Department of Physical Medicine and Rehabilitation, Chung Shan Medical University, Taichun, Taiwan
[5]Department of Medical Research, Chi Mei Medical Center, Tainan, Taiwan
[6]Department of Pediatrics, Chi-Mei Medical Center, Tainan, Taiwan
[7]Department of Pediatrics, Taipei Medical University, Chi Mei Medical Groups, Taipei, Taiwan
[8]Medical School, St George's, University of London, London, United Kingdom
[9]Department of Orthopedics and Traumatology, Taipei Veterans General Hospital, Taipei, Taiwan
[10]School of Medicine, National Yang-Ming University, Taipei, Taiwan
[11]Department of Nursing, College of Medicine, National Cheng Kung University, Tainan, Taiwan
[*]these authors contributed equally

**Corresponding Author:**
Huan-Fang Lee, PhD
Department of Nursing
College of Medicine
National Cheng Kung University
No 988, Chung Hwa Road, Yung Kung District, Tainan 710, Taiwan
Tainan, 710
Taiwan
Phone: 886 +8862812811
Email: Eamonn0330@gmail.com

## Abstract

**Background:** Workplace bullying has been measured in many studies to investigate its effects on mental health issues. However, none have used web-based computerized adaptive testing (CAT) with bully classifications and convolutional neural networks (CNN) for reporting the extent of individual bullying in the workplace.

**Objective:** This study aims to build a model using CNN to develop an app for automatic detection and classification of nurse bullying-levels, incorporated with online Rasch computerized adaptive testing, to help assess nurse bullying at an earlier stage.

**Methods:** We recruited 960 nurses working in a Taiwan Ch-Mei hospital group to fill out the 22-item Negative Acts Questionnaire-Revised (NAQ-R) in August 2012. The k-mean and the CNN were used as unsupervised and supervised learnings, respectively, for: (1) dividing nurses into three classes (n=918, 29, and 13 with suspicious mild, moderate, and severe extent of being bullied, respectively); and (2) building a bully prediction model to estimate 69 different parameters. Finally, data were separated into training and testing sets in a proportion of 70:30, where the former was used to predict the latter. We calculated the sensitivity, specificity, and receiver operating characteristic curve (area under the curve [AUC]), along with the accuracy across studies for comparison. An app predicting the respondent bullying-level was developed, involving the model's 69 estimated parameters and the online Rasch CAT module as a website assessment.

**Results:** We observed that: (1) the 22-item model yields higher accuracy rates for three categories, with an accuracy of 94% for the total 960 cases, and accuracies of 99% (AUC 0.99; 95% CI 0.99-1.00) and 83% (AUC 0.94; 95% CI 0.82-0.99) for the lower and upper groups (cutoff points at 49 and 66 points) based on the 947 cases and 42 cases, respectively; and (2) the 700-case

training set, with 95% accuracy, predicts the 260-case testing set reaching an accuracy of 97. Thus, a NAQ-R app for nurses that predicts bullying-level was successfully developed and demonstrated in this study.

**Conclusions:** The 22-item CNN model, combined with the Rasch online CAT, is recommended for improving the accuracy of the nurse NAQ-R assessment. An app developed for helping nurses self-assess workplace bullying at an early stage is required for application in the future.

***(JMIR Mhealth Uhealth 2020;8(5):e16747)*** doi:10.2196/16747

**KEYWORDS**

nurse bullying; NAQ-R assessment; receiver operating characteristic curve; convolutional neural network; computerized adaptive testing

## Introduction

### Background

Over the past several decades, prevalence rates of workplace bullying have been addressed in a wide range of different studies to investigate bullying's potential effects on mental health [1-4]. Despite all of this attention on workplace bullying, the classification of bullying levels has still not, to this date, reached a consensus in the literature.

### NAQ-R Assessment Used for Examining Workplace Bullying

The 22-item Negative Acts Questionnaire-Revised (NAQ-R) [2,4-7] is one of the most popular tools used for examining individuals who deal with workplace bullying. Using cutting points at –0.7 and 0.7 (or <30 and <60 in the summed score), this test has been proposed to assess nurses to identify their bullying grade from one of three levels (high, moderate, and low) [4]. However, the assessment accuracy for classifying individual bullying levels is challenging and requires improvement due to type I and type II errors.

### Convolutional Neural Networks

Convolutional neural networks (CNNs) have had a significant impact within the field of health informatics [8,9]. Its architecture can be described as an interleaved set of feedforward layers implementing convolutional filters followed by reduction, rectification, or pooling layers [10-12]. For each layer, the CNN creates a high-level abstract feature. The CNN, a famous deep learning method, can improve prediction accuracy up to 7.14% [12] in classification. Accordingly, the 22-item NAQ-R, combined with the CNN technique for improving the prediction accuracy of workplace bullying, is worthy of study.

### Computerized Adaptive Testing With CNN

Computerized adaptive testing (CAT) is based on item response theory (IRT) that adapts to an examinee's ability level [4]. The computer follows an IRT-based algorithm that provides the examinee with the next item, which can be not too hard or not too easy, for answering the next question. As such, each patient needs to answer the fewest possible items, resulting in less respondent burden and even more accurate outcomes [2]. As with all forms of web-based technology development, there has not been an online NAQ-R CAT assessment combined with CNN to assess individual workplace bullying available until now. The issue of missing responses in CAT affecting the CNN computation is one of the problems that limit the development of CAT with CNN. Another limitation for CAT is the numerous parameters within CNN, which are harder to program computer routines than traditional predictive methods, such as multiple regression analysis or logistical regression, which only have a few independent variables in their prediction models.

### Online Assessment Using Smartphones is Required

As the age of digital technology approaches, advances in mobile health (mHealth) and health communication technology are rapidly increasing [13]. Till now, there has been no app for smartphones that measures nurse bullying levels in health care settings. It is not only the complexity of the CAT procedure with multimedia illustrations embedded into the web-based module, but also the difficulty of the model's CNN parameters that need to be transformed into the probability of classification types when the individual bullied-levels are assessed by the NAQ-R CAT. A web-based CAT with CNN app could more accurately alert individuals to alleviate their mental strain before it becomes a serious bullying-victim problem.

### Study Aims

The aims of the current study are to: (1) estimate the model's parameters on the NAQ-R responses by the nurses; and (2) design an app for smartphones based on a website-based assessment of nurse bullying levels.

## Methods

### Data source

The study sample was recruited from three hospitals (Hospital A: 1236-bed medical center; B: 265-bed local hospital; and C: 877-bed region hospital) in southern Taiwan in August 2012. No incentive for participation was offered. A total of 960 copies of the bullying questionnaire were validated, with a return rate of 96.3% [4]. This study was approved and monitored by the Research Ethics Review Board of the Chi-Mei Medical Center. Demographic data were anonymously collected: gender, work tenure in hospitals of all types, age, marital status, and education level.

### Featured Variables

Featured variables include the 22 items in the NAQ-R in which a higher response denotes a more serious bullying problem. The input layer for each case, with 36 elements as a 6×6 image, was constructed with the 22 featured variables and the sequentially repeated responses (eg, the elements from 23 to 36 are followed from the beginning till to the end). The 960 participants were

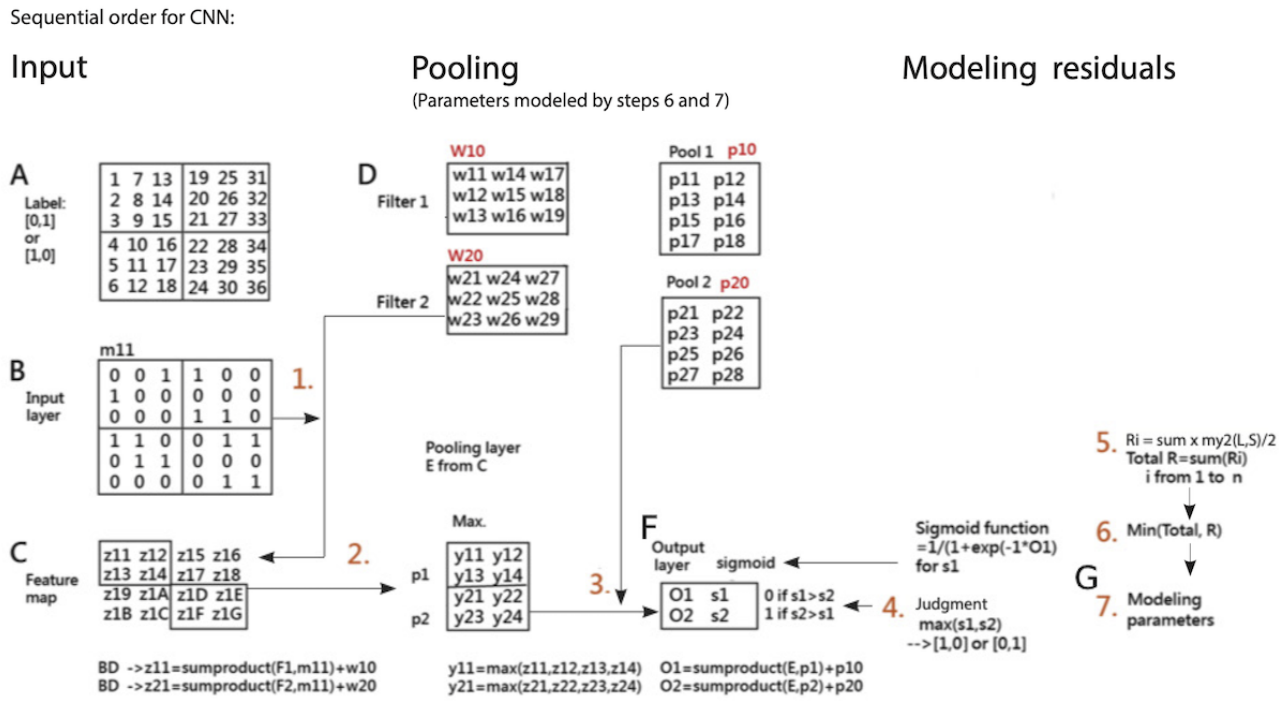
then split into training and testing sets in a proportion of 70:30, where the former was used to predict the latter. The data are shown in Multimedia Appendix 1.

## Unsupervised and Supervised Learnings

Unsupervised learning indicates agnostic aggregation of unlabeled data sets yielding groups or clusters of entities with shared similarities that may be unknown to the user before the analysis step [14,15] (eg, clustering dimensionality reduction using principal component analysis or k-mean clustering). The k-mean clustering aims to partition n observations into k clusters, in which each observation belongs to the cluster with the nearest mean [16]. Two sets of two and three categories each were clustered in comparison to this study. In contrast, supervised learning employs "labeled" training data sets (defined by the previous approach of k-mean clustering) to yield a qualitative or quantitative output through the CNN algorithm [14,17].

In this study, the k-mean was used as unsupervised learning for: (1) clustering participants into two classes (eg, the three categories of suspicious mild [n=918], moderate [n=29], and severe [n=13]). CNN was applied as supervised learning to build a bully prediction model for estimating the 69 parameters. See Figure 1 and 2 for more detailed information.

**Figure 1.** Interpretation of the CNN algorithm in Microsoft Excel. CNN: convolutional neural network.




## CNN Applied to This Study

CNN is a variant of the standard multilayer perceptron, and it is especially used for pattern recognition compared with conventional approaches [18] due to its capability to reduce the dimension of data, extract a feature sequentially, and classify one structure of the network [19]. The basic CNN model was inspired in 1962, from the visual cortex proposed by Hubel and Wiesel [18]. To simplify the CNN concept and process, we present it in Figure 1 (see Multimedia Appendix 2 for more detailed information on interpretation).

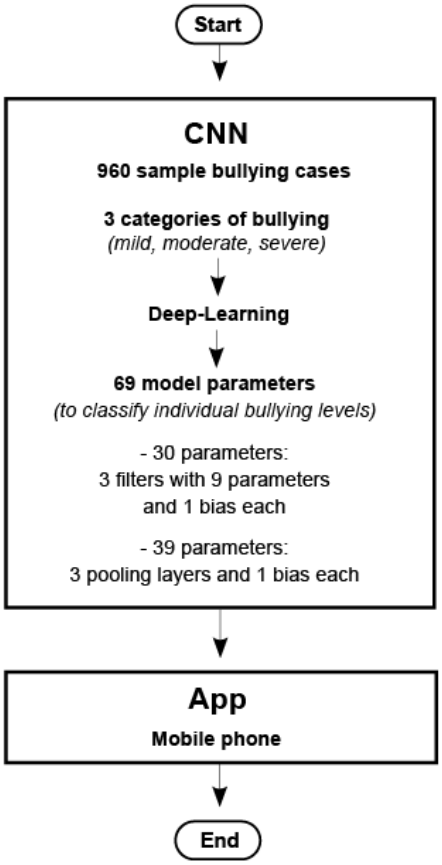
## Tasks for Performing CNN

Task 1 is the comparison of prediction accuracies in the tree-category model. Two sets of categories (ie, 2 and 3) on 960 cases were mirrored to compare, first, the prediction accuracies (eg, the sensitivity, specificity, receiver operating characteristic (ROC) curve, and area under the curve [AUC]) using k-mean clustering. Task 2 is validation compared to the training and testing sets. We used the known responses and their corresponding labels (ie, suspicion of bullying levels) to build a model for predicting the unknown label of the specific responses. The 960 cases were split into training and testing sets in a proportion of 70:30, with the former used to predict the latter. The accuracy rates in these two sets were compared. Finally, task 3 is the app detecting bullied levels for a web-based assessment. A 22-item self-assessment app using mobile phones was designed to predict bullying levels using the CNN algorithm and the model parameters [20]. The resulting classification was based on the 22-item model.

## Statistical Tools and Data Analysis

MedCalc 9.5.0.0 for Windows (MedCalc Software, Ostend, Belgium) was used to calculate the sensitivity, specificity, and corresponding AUC using logistic regression when the observed labels and the predicted probabilities (ie, the a2 calculated by the sigmoid function in the output layer) were applied. A visual representation displaying the classification effect was plotted using the Rasch category characteristic curve (CCC) [21,22]. The study flowchart and the CNN modeling process are shown in Figure 2 and Multimedia Appendix 2, respectively.

**Figure 2.** The study flowchart. CNN: convolutional neural network.




## Results

### Demographic Data of the 960 Cases

A sample of 960 nurses was obtained for the study. The mean age of the participants was 32.7 (SD 5.8) years old, 96% (n=922) were female, and more than 57.5% (n=553) were unmarried (Table 1).

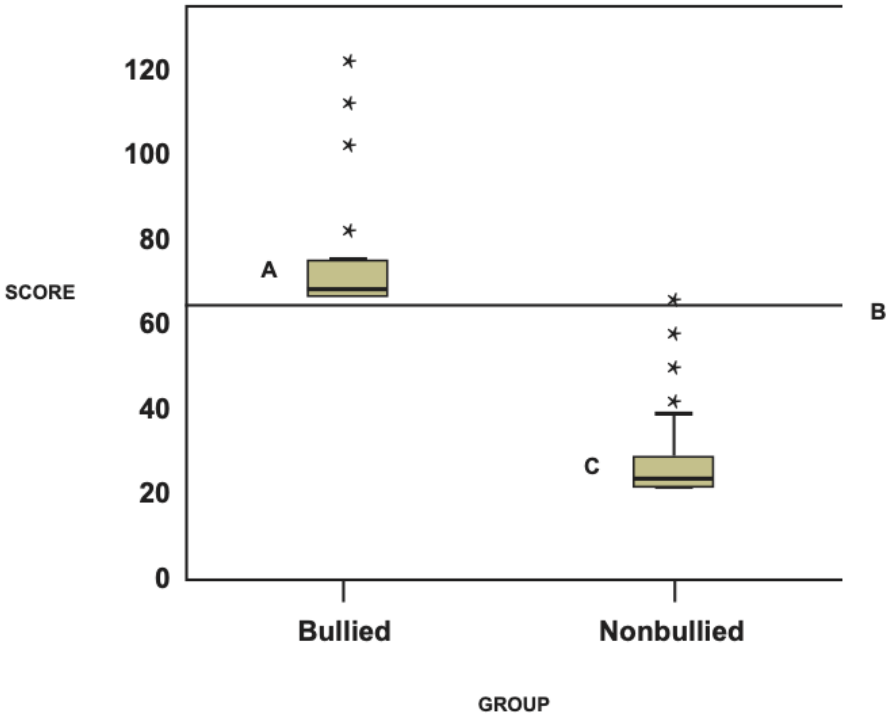
**Table 1.** Demographic data of the study sample.

| Variables | n (%) |
|---|---|
| **Hospital** | |
| Hospital A | 542 (56.4) |
| Hospital B | 323 (33.6) |
| Hospital C | 95 (10) |
| **Gender** | |
| Male | 38 (4) |
| Female | 922 (96) |
| **Education** | |
| High school | 6 (0.6) |
| College | 464 (48.3) |
| University | 474 (49.3) |
| Graduate school | 16 (1.8) |
| **Marriage** | |
| Unmarried | 553 (57.5) |
| Married | 403 (42.1) |
| Divorced | 4 (0.4) |
| **Nursing grade** | |
| N0 | 34 (3.5) |
| N1 | 281 (29.3) |
| N2 | 316 (32.9) |
| N3 | 243 (25.3) |
| N4 | 86 (8.9) |
| **Title** | |
| Nurse | 772 (80.3) |
| Chief | 158 (17.7) |
| Leader | 8 (0.8) |
| Others | 12 (1.2) |

## Task 1: Comparison of Prediction Accuracies in the Tree-Category Model

Two groups divided by the k-mean clustering are shown in Figure 3. The cuttoff point is set at 66 points. Another visual representation displaying the classification effect is plotted using a box plot (see Figure 3).

**Figure 3.** Two groups divided by the k-mean clustering. A) n=35; B) Cuttoff point set at 66 points; C) n=925.




We can see that the 22-item model yields higher accuracy rates for three categories, with an accuracy of 94% for the total 960 cases, and accuracies of 99% (AUC 0.99; 95% CI 0.99-1.00) and 83% (AUC 0.94; 95% CI 0.82-0.99) for the lower and upper groups (cutoff points at 49 and 66) based on the 947 cases and 42 cases, respectively (see Figure 4 and Tables 2 and 3).

**Figure 4.** The bullied classes clustered with 3 categories using cut-off points to identify the sensitivity and specificity with AUC (area under the curve).

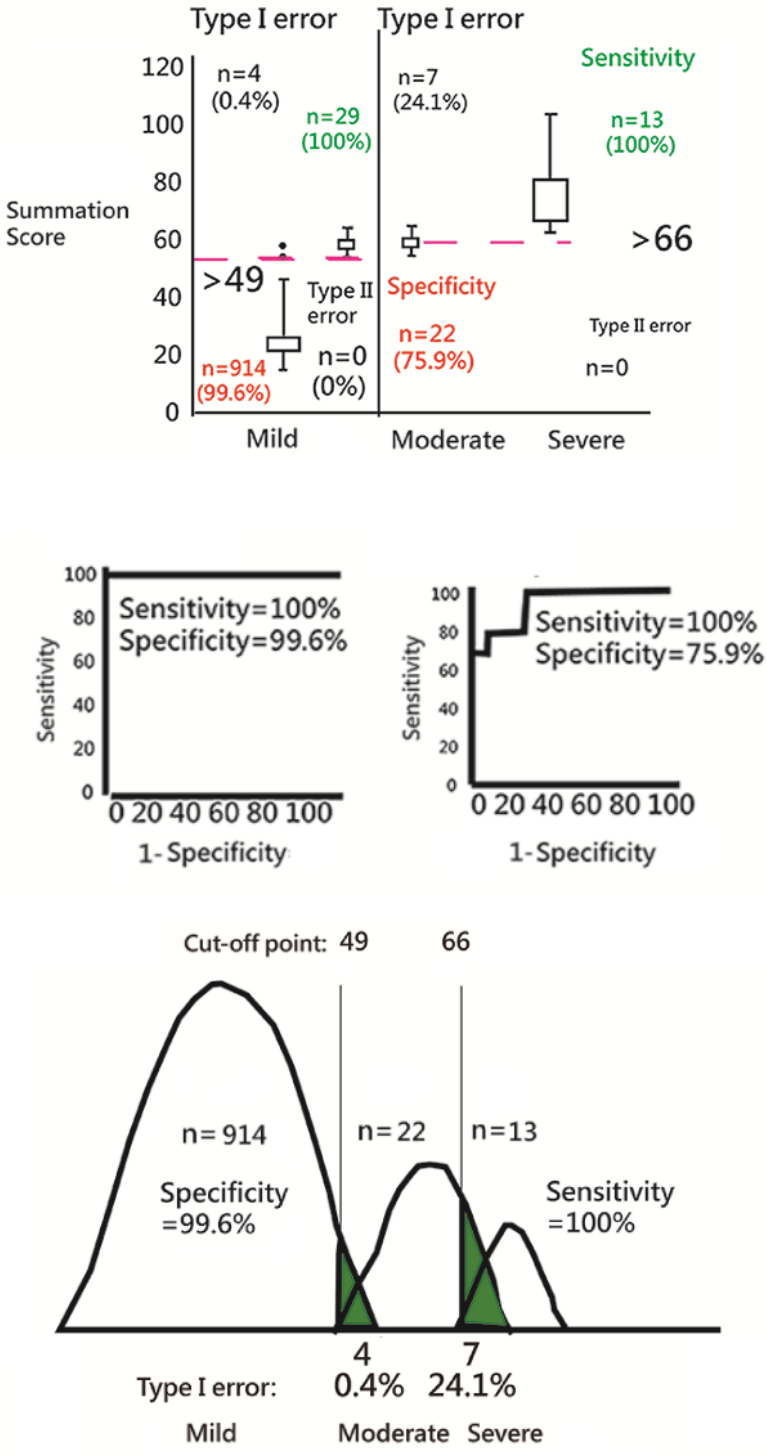


**Table 2.** Mild and moderate scenario applied to CNN for the prediction of nurse bullying levels.

| | True condition | | | |
|---|---|---|---|---|
| Scenario A (22 items), ACC[a]=0.99, (n=947) | Positive | Negative | PPV[b]/FOR[c] | FDR[d]/NPV[e] |
| Positive | 29 | 4 | 0.88 | 0.12 |
| Negative | 0 | 914 | 0 | 1 |
| Sensitivity | 1 | __[f] | — | — |
| FPR[g] | 0.01 | — | — | — |
| FNR[h] (Miss rate) | 0 | — | — | — |
| Specificity | 0.99 | — | — | — |
| AUROC[i] (95% CI) | 0.99 (0.99-1) | — | — | — |

[a]ACC: accuracy
[b]PPV: positive predictive value.
[c]FOR: 1-PPV.
[d]FDR: 1-NPV.
[e]NPV: negative predictive value.
[f]Not applicable.
[g]FPR: false positive rate.
[h]FNR: false negative rate.
[i]AUROC: area under the receiver operating characteristic curve.

**Table 3.** Moderate and severe scenario applied to CNN for the prediction of nurse bullying levels.

| | True condition | | | |
|---|---|---|---|---|
| Scenario B (22 items), ACC[a]=0.83, (n=42) | Positive | Negative | PPV[b]/FOR[c] | FDR[d]/NPV[e] |
| Positive | 13 | 7 | 0.65 | 0.35 |
| Negative | 0 | 22 | 0 | 1 |
| Sensitivity | 1 | __[f] | — | — |
| FPR[g] | 0.24 | — | — | — |
| FNR[h] (Miss rate) | 0 | — | — | — |
| Specificity | 0.76 | — | — | — |
| AUROC[i] (95% CI) | 0.94 (0.82-0.99) | — | — | — |

[a]ACC: accuracy
[b]PPV: positive predictive value.
[c]FOR: 1-PPV.
[d]FDR: 1-NPV.
[e]NPV: negative predictive value.
[f]Not applicable.
[g]FPR: false positive rate.
[h]FNR: false negative rate.
[i]AUROC: area under the receiver operating characteristic curve.

### Task 2: Validation Compared to the Training and Testing Sets

The 700-case training set with an accuracy of 95% predicts the 260-case testing set reaching an accuracy of 97%. Interested readers are encouraged to see the study process in Multimedia Appendix 2, using the parameters modeled by the 700-case training set to predict the accuracy in the 260-case testing set.

### Task 3: App Detecting the Bullied Levels on a Web-Based Assessment

A NAQ-R app for nurses predicting individual bullying levels was developed and demonstrated in Figure 5. One resulting example of the mild level is present at the bottom in Figure 5 on the CCC (ie, category 0 from the left-top to the right-bottom corner, category 1 in the middle, and category 2 from the left-bottom to the top-right side) based on the Rasch rating scale

model [21,22], which is novel when using a visual display shown on Google Maps.

Interested readers are invited to scan the QR code to practice the NAQ-R app on their own. It is worth noting that all 69 model parameters for classifying individual bullying levels are involved in the Rasch online CAT module.

**Figure 5.** Snapshots on a mobile phone responding to questions (top) and the result (bottom) for assessing individual bullied levels.

## Discussion

### Principal Findings

We observed that: (1) the 22-item model for three categories yields higher accuracy rates; and (2) the 700-case training set with an accuracy of 95% predicts the 260-case testing set reaching an accuracy of 97%. We also developed and demonstrated an NAQ-R app for nurses that predicts bullying level in this study.

The difference between the traditional score calculation method and the new model using CNN can be described as: the traditional score calculation lacks the bullying classification. As such, cutoff points are the way to classify the extent of bullying at the workplace. Nonetheless, the cutoff points drive type I (ie, false negative) and type II (ie, false positive) errors higher than the CNN algorithm.

The app created to examine and the extent of workplace bullying for individuals has two parts: (1) the Rasch CAT; and (2) classification using CNN. However, not all items answered in the Rasch CAT results were missing responses on CNN. We thus applied the Rasch rating scale model [21,22] for generating the expected responses and overcame the drawback of not having all the items answered in the CAT.

### What This Knowledge Adds to What We Already Know

The NAQ-R app has been the most widely used tool for measuring workplace bullying in the world [2,4-7]. Over 32 articles were found by searching the keywords “NAQ-R” as of September 30, 2019. However, none provided an acceptable scheme to classify the individual bullying levels (ie, mild, moderate, and severe). The previous study [4] provided a cutoff point scheme (ie, −0.7 and 0.7 logits using CAT to measure the extent of bullying) and claimed the prevalence rate of bullying for nurses was 1.5%. In this study, the cutoff points for three categories (ie, mild, moderate, and severe) are set at >49 and >66 when the total score is 110 using k-mean clustering, which are different from those set at <30 and <60 in the summed score [4] on the assumption of an equal sample size across the levels (ie, mild, moderate, and severe).

However, no matter which cutoff point scheme is applied (eg, Figure 4), misclassifications must exist due to their Type I ($\alpha$) and II (1-$\beta$) errors [23]. In contrast, the CNN model can minimize Type I and II errors and improve the prediction accuracy (up to 7.14 %) [12], which is one of the features of this study.

### What it Implies and What Should be Changed

Not all questions were answered in CAT. Different from those using the mean value [9] over the entire dataset to fill the missing values, we applied the expected value in the model for each unanswered response to fill the missing data, as done in previous studies [24,25]. By doing so, the expected responses and the CNN parameters can thus be applied to classify the groups of individual bullying levels. So far, we have not seen anyone using the CNN approach to predict nurse bullying levels in the literature, which is a breakthrough, and the second feature of this study.

Over 708 articles were found using the keyword “convolutional neural network” (Title) when searching in PubMed Central as of September 23, 2019. None of the studies found used Excel (Microsoft Corp, Redmond, Washington, United States) to perform the CNN. The interpretations of the CNN concept and the process, and the parameter estimations, are shown in Figure 1 and Multimedia Appendices 2-4, which is the third feature of this study.

Furthermore, at the end of 2019, 200 papers were collected from the US National Library of Medicine National Institutes of Health when searching the keywords “computer adaptive testing.” None that were published used an online assessment with CNN suited for smartphones, and thus were not applicable for this study. We believe that more papers in the future will be published on the usefulness of online CAT with CNN, because all forms of web-based technology are rapidly increasing [13], so a need for classification assessment in clinical settings will also increase.

### Strengths of This Study

It is easy to set up an online CAT assessment form if the designer uploads relevant audio and visual files to the corresponding questions in the database. We applied the CNN algorithm along with the model’s parameters to design the routine on an app that is used to detect individual bullying levels for nurses in hospitals (see Figure 5), which is the fifth feature of this study. We have not seen any such NAQ-R [2,4-7] CAT combined with CNN implemented on mobile phones before.

As with all forms of web-based technology, advances in mHealth and health communication technology are rapidly emerging [13]. Mobile online CAT assessment is promising and worth considering in many fields of health assessment. An online CAT assessment, such as the one we developed, can be used to inform examinees quickly about when and whether they should take actions or follow-up with a psychiatrist, and how to improve their behaviors and attitudes given that their lifestyle is not changed. Mobile online CAT assessment is promising, and is worth using it to promote nurses’ health literacy. It is recommended that interested readers scan the QR codes on Figure 5, one for the app and another for the MP4, and see: (1) the details about responding to questions; or (2) the real experience of answering the 22-item NAQ-R CAT with the CNN algorithm for a website assessment.

### Limitations and Future Studies

Our study has some limitations. First, although the psychometric properties of the 22-item NAQ-R have been validated for measuring workplace bullying [2,4-7], there is no evidence to support that the 22-item NAQ-R is suitable for use on CAT assessment. We recommend additional studies using their own k-mean algorithm and CNN model to estimate the parameters and see whether a difference exists. Second, although the three classes were determined by k-mean clustering with the CNN algorithm, which can increase accuracy rates (see Tables 2 and 3), we cannot guarantee that this CNN is the only thing improving classification accuracy. Future studies are encouraged

to look for other types of prediction methods that can also improve the power of the model prediction, such as Logistic regression, Naïve Bayes, Decision trees, Random Forests, and Gradient tree boosting [26-35]. They could also use other artificial neural networks, such as a Feedforward Neural Network, a Radial Basis Function Neural Network, a Multilayer Perceptron, a Recurrent Neural Network, a Modular Neural Network, or a Sequence-To-Sequence Model [36]. Third, the study was based on publications [2,4] that used the 22-item NAQ-R CAT module. All the model parameters (ie, item difficulties and step-threshold difficulties) were derived from those studies. If any environment or condition is changed (eg, for other professionals or workplaces), the result (eg, the model's parameters) will be different from the current study and worth verifying in the future. Fourth, the NAQ-R is a one-dimensional construct. The item difficulties used to estimate the person measure were calibrated by using the Rasch Winsteps software (Winsteps.com, Chicago, United States). A person's ability (θ) should be further estimated by the computer adaptive testing method [2,4]. Similarly, a person's ability (θ) should be known when the respondent completes the NAQ-R CAT on an app. Otherwise, the remaining items that were not answered in the CAT could not be computed for the website assessment that is used to obtain the expected responses and classify the bullying levels using the CNN algorithm. Future studies should be cautious about this matter. Fifth, the way to access the app via scanning the QR code in Figure 5. the professionally practical app should be further developed for android and IOS in the future. Finally, the study sample was taken from a nurse survey. The model parameters estimated for the NAQ-R are suitable for professionals and the workplace, but generalizing these workplace bullying assessment findings (eg, the cutoff points; see Figure 4) might be somewhat limited because the sample consisted only of nurses working at hospitals. Additional studies are needed to reexamine whether the psychometric properties of the workplace bullying assessment are like those of other worksites in/out of a hospital.

## Conclusion

The contributions in this study include: (1) overcoming the problem of missing responses that affects CNN computation and limits CAT development combined with the CNN; (2) introducing CNN availability in Microsoft Excel; (3) demonstrating an app that incorporates Rasch CAT with numerous parameters in CNN. The 22-item NAQ-R CAT is recommended for combining the parameters estimated in CNN to improve the accuracy of determining individual bullying levels. An app developed for helping nurses' self-assess workplace bullying is at an early stage but is required for application in the future.

## Authors' Contributions

SC conceived and designed the study. YT and PH performed the statistical analyses. JCC was in charge of recruiting study participants. WC and TWC helped design the study, collected information, and interpreted data. HF monitored the research. All authors read and approved the final article.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Study dataset (MS Excel).
[XLSX File (Microsoft Excel File), 2117 KB - mhealth_v8i5e16747_app1.xlsx ]

Multimedia Appendix 2
CNN using MS Excel to interpret on Figure 1.
[DOCX File , 893 KB - mhealth_v8i5e16747_app2.docx ]

Multimedia Appendix 3
CNN performed in Excel.
[DOCX File , 13 KB - mhealth_v8i5e16747_app3.docx ]

Multimedia Appendix 4
App Online assessing nurse workplace bullying.
[DOCX File , 13 KB - mhealth_v8i5e16747_app4.docx ]

## References

1. Nielsen M, Notelaers G, Einarsen S. Measuring exposure to workplace bullying. In: Ininarsen S, Hoel H, Zapf D, Cooper CL, editors. Bullying and harassment in the workplace: developments in theory, research, and practice. Boca Raton, Florida: CRC Press; 2011.

2. Ma S, Chien T, Wang H, Li Y, Yui M. Applying computerized adaptive testing to the Negative Acts Questionnaire-Revised: Rasch analysis of workplace bullying. J Med Internet Res 2014 Feb 17;16(2):e50 [FREE Full text] [doi: 10.2196/jmir.2819] [Medline: 24534113]
3. Abe K, Henly SJ. Bullying (Ijime) Among Japanese Hospital Nurses. Nursing Research 2010;59(2):110-118. [doi: 10.1097/nnr.0b013e3181d1a709]
4. Ma S, Wang H, Chien T. A new technique to measure online bullying: online computerized adaptive testing. Ann Gen Psychiatry 2017 Jul 3;16(1):26 [FREE Full text] [doi: 10.1186/s12991-017-0149-z] [Medline: 28680455]
5. Makarem NN, Tavitian-Elmadjian LR, Brome D, Hamadeh GN, Einarsen S. Assessment of workplace bullying: reliability and validity of an Arabic version of the Negative Acts Questionnaire-Revised (NAQ-R). BMJ Open 2018 Dec 09;8(12):e024009 [FREE Full text] [doi: 10.1136/bmjopen-2018-024009] [Medline: 30530587]
6. Fullerton L, Oglesbee S, Weiss SJ, Ernst AA, Mesic V. Assessing the Prevalence and Predictors of Bullying Among Emergency Medical Service Providers. Prehosp Emerg Care 2018 Aug 17;23(1):1-6. [doi: 10.1080/10903127.2018.1470208] [Medline: 30118605]
7. El Ghaziri M, Storr CL, Simons SR, Trinkoff AM, McPhaul KM, London M, et al. Comparative psychometric review of the Negative Acts Questionnaire-Revised in a unionized U.S. public sector workforce. WOR 2019 Feb 07;62(1):161-171. [doi: 10.3233/wor-182851]
8. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. IEEE J. Biomed. Health Inform 2017 Jan;21(1):4-21. [doi: 10.1109/jbhi.2016.2636665]
9. Wang H, Cui Z, Chen Y, Avidan M, Abdallah AB, Kronzer A. Predicting Hospital Readmission via Cost-Sensitive Deep Learning. IEEE/ACM Trans. Comput. Biol. and Bioinf 2018 Nov 1;15(6):1968-1978. [doi: 10.1109/tcbb.2018.2827029]
10. Tobore I, Li J, Yuhang L, Al-Handarish Y, Kandwal A, Nie Z, et al. Deep Learning Intervention for Health Care Challenges: Some Biomedical Domain Considerations. JMIR Mhealth Uhealth 2019 Aug 02;7(8):e11966 [FREE Full text] [doi: 10.2196/11966] [Medline: 31376272]
11. Kwon S, Hong J, Choi E, Lee E, Hostallero DE, Kang WJ, et al. Deep Learning Approaches to Detect Atrial Fibrillation Using Photoplethysmographic Signals: Algorithms Development Study. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e12770 [FREE Full text] [doi: 10.2196/12770] [Medline: 31199302]
12. Sathyanarayana A, Joty S, Fernandez-Luque L, Ofli F, Srivastava J, Elmagarmid A, et al. Sleep Quality Prediction From Wearable Data Using Deep Learning. JMIR Mhealth Uhealth 2016 Nov 04;4(4):e125. [doi: 10.2196/mhealth.6562]
13. Mitchell SJ, Godoy L, Shabazz K, Horn IB. Internet and mobile technology use among urban African American parents: survey study of a clinical population. J Med Internet Res 2014 Jan 13;16(1):e9 [FREE Full text] [doi: 10.2196/jmir.2673] [Medline: 24418967]
14. Rashidi HH, Tran NK, Betts EV, Howell LP, Green R. Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. Acad Pathol 2019 Sep 03;6:2374289519873088 [FREE Full text] [doi: 10.1177/2374289519873088] [Medline: 31523704]
15. Buehler L, Rashidi H. Bioinformatics Basics, Application in Biological Science and Medicine. Boca Raton, Florida: CRC Press; 2005.
16. Chen C, Luo J, Parker K. Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications. IEEE Trans Image Process 1998 Feb 2;7(12):1673-1683. [doi: 10.1109/83.730379] [Medline: 18276234]
17. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. 2006 Presented at: Proceedings of the 23rd International Conference on Machine learning; 25-29 June; Pittsburgh, Pennsylvania. [doi: 10.1145/1143844.1143865]
18. Guzmán MG, Kouri G. Dengue: an update. The Lancet Infectious Diseases 2002 Jan;2(1):33-42. [doi: 10.1016/s1473-3099(01)00171-2]
19. Bengio Y. Learning Deep Architectures for AI. FNT in Machine Learning 2009;2(1):1-127. [doi: 10.1561/2200000006]
20. Saha S. towardsdatascience.com. 2018 Dec 15. A Comprehensive Guide to Convolutional Neural Networks — the ELI5 way URL: https://tinyurl.com/y9mmosuq [accessed 2020-02-11]
21. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago, Illinois: University of Chicago Press; 1980.
22. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978 Dec;43(4):561-573. [doi: 10.1007/bf02293814]
23. Chien T, Lin W. Improving Inpatient Surveys: Web-Based Computer Adaptive Testing Accessed via Mobile Phone QR Codes. JMIR Med Inform 2016 Mar 02;4(1):e8 [FREE Full text] [doi: 10.2196/medinform.4313] [Medline: 26935793]
24. Lee Y, Chou W, Chien T, Chiu P, Yeh Y, Lee H. An App Developed for Detecting Nurse Burnouts Using the Convolutional Neural Networks in Microsoft Excel. JMIR Medical Informatics 2020 (forthcoming). [doi: 10.2196/16528]
25. Chien T, Lin W. Simulation study of activities of daily living functions using online computerized adaptive testing. BMC Med Inform Decis Mak 2016 Oct 10;16(1):130 [FREE Full text] [doi: 10.1186/s12911-016-0370-8] [Medline: 27724939]
26. Nelder JA, Wedderburn RWM. Generalized Linear Models. Journal of the Royal Statistical Society. Series A (General) 1972;135(3):370 [FREE Full text] [doi: 10.2307/2344614]

27. Cox DR. The Regression Analysis of Binary Sequences. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;20(2):215-232. [doi: 10.1111/j.2517-6161.1958.tb00292.x]
28. Szumilas M. Explaining odds ratios. J Can Acad Child Adolesc Psychiatry 2010 Aug;19(3):227-229 [FREE Full text] [Medline: 20842279]
29. Bayes T, Price R.. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Phil. Trans. R. Soc 1997 Jan;53:370-418 [FREE Full text] [doi: 10.1098/rstl.1763.0053]
30. Domingos P, Pazzani M. On the Optimality of the Simple Bayesian Classifier under Zero-One Loss. Mach Learn 1997;29:103-130 [FREE Full text]
31. Quinlan JR. Induction of decision trees. Mach Learn 1986 Mar;1(1):81-106. [doi: 10.1007/bf00116251]
32. Breiman L. Random Forests. In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning and Data Mining. Boston, Massachusetts: Springer; 2017:32.
33. Tin KH. Random decision forests. 1995 Presented at: Proceedings of 3rd International Conference on Document Analysis and Recognition; 14-16 Aug; Montreal, Canada. [doi: 10.1109/icdar.1995.598994]
34. Friedman JH. Stochastic gradient boosting. Computational Statistics & Data Analysis 2002 Feb;38(4):367-378. [doi: 10.1016/s0167-9473(01)00065-2]
35. Freund Y, Mason L. The Alternating Decision Tree Algorithm. 1999 Presented at: Proceedings of the 16th International Conference on Machine Learning; 26-27 June; Ithaca, New York URL: https://cseweb.ucsd.edu/~yfreund/papers/atrees.pdf
36. Mehta A. Digital Vidya. 2019 Jan 25. A Comprehensive Guide to Types of Neural Networks URL: https://www.digitalvidya.com/blog/types-of-neural-networks/ [accessed 2020-02-12]


## Abbreviations

**AUC:** area under the curve
**CAT:** computerized adaptive testing
**CCC:** category characteristic curve
**CNN:** convolutional neural network
**IRT:** item response theory
**mHealth:** mobile health
**NAQ-R:** the 22-item Negative Acts Questionnaire-Revised
**ROC:** receiver operating characteristic


*Edited by G Eysenbach; submitted 21.10.19; peer-reviewed by L Shen, S Chen; comments to author 23.12.19; revised version received 02.01.20; accepted 26.01.20; published 20.05.20.*

*Please cite as:*
*Ma SC, Chou W, Chien TW, Chow JC, Yeh YT, Chou PH, Lee HF*
*An App for Detecting Bullying of Nurses Using Convolutional Neural Networks and Web-Based Computerized Adaptive Testing: Development and Usability Study*
*JMIR Mhealth Uhealth 2020;8(5):e16747*
*URL: https://mhealth.jmir.org/2020/5/e16747*
*doi:10.2196/16747*
*PMID:32432557*

Original Paper

# A Mobile Phone–Based Sexual and Reproductive Health Intervention for Female Sex Workers in Kenya: Development and Qualitative Study

Frances H Ampt[1,2*], MPH; Kelly L'Engle[3*], PhD; Megan S C Lim[1,2], PhD; Kate F Plourde[4], MPH; Emily Mangone[5], MSc; Collins Mudogo Mukanya[6], MA; Peter Gichangi[7,8,9], PhD; Griffins Manguro[9], MPH; Margaret Hellard[1,10], PhD; Mark Stoové[1,2], PhD; Matthew F Chersich[8,11], PhD; Walter Jaoko[12], PhD; Paul A Agius[1,2], MSc; Marleen Temmerman[8,9,13], PhD; Winnie Wangari[9], BA; Stanley Luchters[1,2,8,13], PhD

[1]Burnet Institute, Melbourne, Australia
[2]Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia
[3]University of San Francisco, San Francisco, CA, United States
[4]FHI 360, Durham, NC, United States
[5]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[6]mHealth Kenya, Nairobi, Kenya
[7]Technical University of Mombasa, Mombasa, Kenya
[8]Department of Public Health and Primary Care, International Centre for Reproductive Health, Ghent University, Ghent, Belgium
[9]International Centre for Reproductive Health, Mombasa, Kenya
[10]Department of Infectious Diseases, The Alfred Hospital, Melbourne, Australia
[11]Wits Reproductive Health and HIV Institute, University of the Witwatersrand, Johannesburg, South Africa
[12]University of Nairobi, Nairobi, Kenya
[13]Aga Khan University, Nairobi, Kenya
[*]these authors contributed equally

**Corresponding Author:**
Frances H Ampt, MPH
Burnet Institute
85 Commercial Rd
Melbourne, 3004
Australia
Phone: 61 2 9282 2111
Email: frances.ampt@burnet.edu.au

## Abstract

**Background:** Female sex workers (FSWs) have high rates of both unintended pregnancy and HIV, but few health promotion interventions address their contraceptive needs or other sexual and reproductive health and rights (SRHR) concerns. A broader approach integrates contraceptive promotion with HIV and sexually transmitted infection (STI) prevention and management, alcohol awareness, gender-based violence and rights, and health care utilization. The Women's Health Intervention using SMS for Preventing Pregnancy (WHISPER) mobile phone intervention uses a participatory development approach and behavior change theory to address these high-priority concerns of FSWs in Mombasa, Kenya.

**Objective:** This paper aimed to (1) describe the process of development of the WHISPER intervention, its theoretical framework, key content domains and strategies and (2) explore workshop participants' responses to the proposed intervention, particularly with regard to message content, behavior change constructs, and feasibility and acceptability.

**Methods:** The research team worked closely with FSWs in two phases of intervention development. First, we drafted content for three different types of messages based on a review of the literature and behavior change theories. Second, we piloted the intervention by conducting six workshops with 42 FSWs to test and refine message content and 12 interviews to assess the technical performance of the intervention. Workshop data were thematically analyzed using a mixed deductive and inductive approach.

**Results:** The intervention framework specified six SRHR domains that were viewed as highly relevant by FSWs. Reactions to intervention content revealed that social cognitive strategies to improve knowledge, outcome expectations, skills, and self-efficacy resonated well with workshop participants. Participants found the content empowering, and most said they would share the messages with others. The refined intervention was a 12-month SMS program consisting of informational and motivational messages, role model stories portraying behavior change among FSWs, and on-demand contraceptive information.

**Conclusions:** Our results highlight the need for health promotion interventions that incorporate broader components of SRHR, not only HIV prevention. Using a theory-based, participatory approach, we developed a digital health intervention that reflects the complex reality of FSWs' lives and provides a feasible, acceptable approach for addressing SRHR concerns and needs. FSWs may benefit from health promotion interventions that provide relevant, actionable, and engaging content to support behavior change.

***(JMIR Mhealth Uhealth 2020;8(5):e15096)*** doi:10.2196/15096

**KEYWORDS**

sex work; mobile health (mHealth); unintended pregnancy; qualitative research

## Introduction

HIV prevention programs for female sex workers (FSWs) utilizing peer educators, drop-in-centers, and mobile outreach have been implemented in sub-Saharan Africa [1,2] and have shown promise in improving condom use, HIV and sexually transmitted infection (STI) prevalence, and HIV testing [3-5]. However, the broader sexual and reproductive health and rights (SRHR) needs of this population have been largely neglected by a narrow focus on HIV [6], potentially limiting the effectiveness of prevention programs [7] and prompting calls for greater integration of family planning, community empowerment, gender-based violence, and antenatal care services into existing programs [1,8-12].

Pregnancy prevention is a particular area of need for FSWs, with high rates of unintended pregnancy and low uptake of highly effective contraception and dual method use among those wanting to avoid pregnancy [13,14]. Research with FSWs in Mombasa, a port city and transport hub on Kenya's East Coast with a large FSW population [15], documented that over 1 year, 24% had an unintended pregnancy and only 57% were using a modern contraceptive method [16].

Limited knowledge of long-acting reversible contraceptives (LARCs), fear of side effects, and social and gender norms that limit the use of family planning are common among FSWs in this setting [11,16-19] and women in sub-Saharan Africa more generally [20,21]. This indicates a critical need for messaging that addresses family planning knowledge, attitudes, and behaviors in the context of sex work. Mobile phones offer a promising medium for such communication, as they are increasingly used to arrange sex work encounters and solicit clients [22], can reach marginalized populations with low engagement in formal services, and mobile coverage is high in most countries (eg, 96% in Kenya) [23].

Mobile phones have been used to deliver health promotion in a variety of contexts, and this approach has been effective in improving knowledge, use, and continuation of contraception [24], as well as impacting preventive behaviors for other health domains [25]. Mobile health (mHealth) interventions have not been implemented with FSWs, but they have been evaluated with young people and postpartum women in sub-Saharan Africa, and have successfully impacted contraceptive outcomes in these contexts [26-28].

We developed a mobile phone intervention for FSWs in Mombasa to promote contraceptive use—particularly LARCs—and other behaviors related to SRHR. This intervention, called the Women's Health Intervention using SMS for Preventing Pregnancy (WHISPER), is being tested in a cluster-randomized controlled trial (RCT) to assess its impact on unintended pregnancy [29].

The intervention was developed using a participatory design approach. FSWs in Mombasa were involved in the initial conception of the intervention and in formal workshopping and testing. Participation by the target community in intervention design [30] and the development of health programs [31] may lead to greater health impacts. However, participatory design methods for mHealth interventions with minority populations such as FSWs are rarely explicitly described [32].

In this paper, we aim to (1) describe the development of the WHISPER intervention and present its theoretical framework, key content domains, and strategies, and (2) explore workshop participants' responses to the proposed intervention, particularly with regard to message content and behavior change constructs. Finally, we present the schedule and approach for intervention implementation and delivery.

## Methods

### Summary

Methods for the development of WHISPER have been described by Ampt et al [29] and generally follow the steps outlined by L'Engle et al [33]. The intervention was developed in two phases: first, to design the intervention framework and draft content; and second, to pilot the intervention with FSWs, refine the messages based on the results, and finalize the intervention structure and content.

### Phase 1: Developing the Framework and Draft Content

The framework for intervention content, and the drafting of initial messages, was informed by the following: review of the literature on motivators and barriers to FSWs' adoption of healthy SRHR behaviors; consideration of health promotion

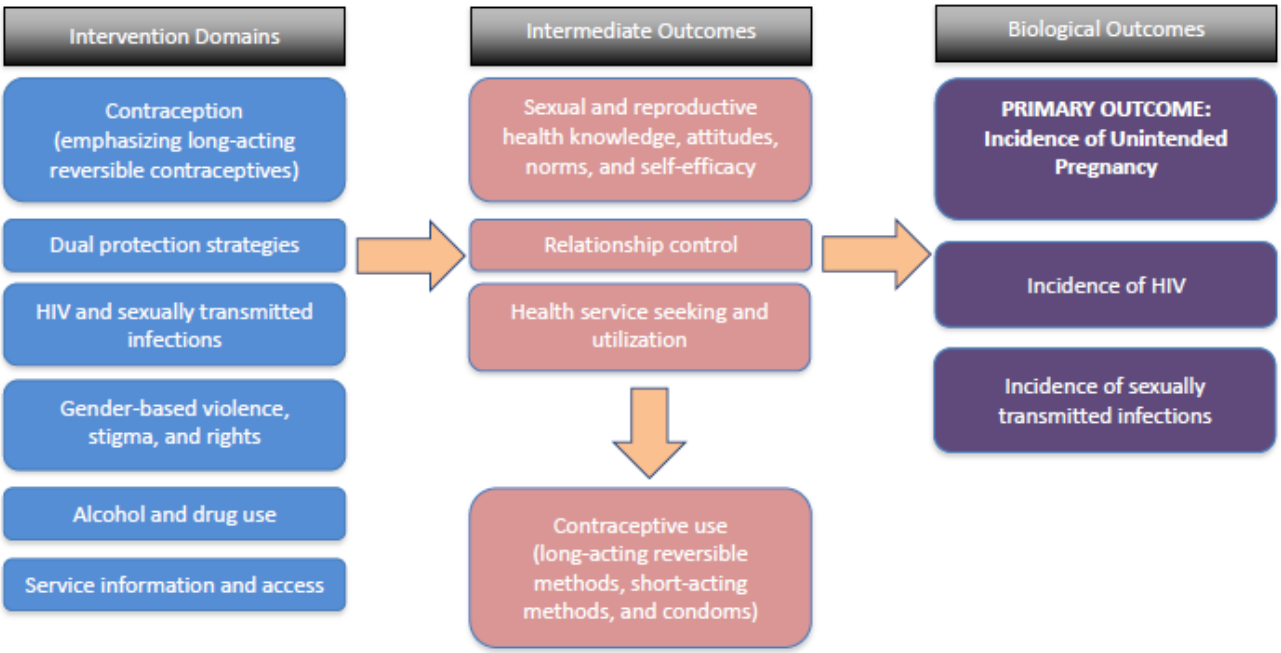
theory, specifically transtheoretical [34] and social cognitive [35] theories; and consultation with FSWs who formed part of the research team. These women were experienced peer educators at the International Center for Reproductive Health's drop-in centers and came from the targeted FSW communities. We incorporated messaging from existing mHealth repositories and previous programs developed by the investigators [36-38] and aligned the content with relevant Kenyan and global guidelines for family planning [39,40] tailored to the specific needs of sex workers including their high risk of STIs and HIV [41]. We drafted and tested the messages in English rather than Kiswahili following advice from the Kenyan research team and peer educators.

### *Intervention Framework*

A review of the literature and behavior change theory highlighted key content domains and corresponding behavioral factors that impact the risk of unintended pregnancy, STIs, and HIV.

These domains and factors were confirmed as important and relevant to FSWs during consultations with peer educators and were incorporated into a logic model (Figure 1) that guided content development. The peer educators agreed that pregnancy prevention was a high priority and also identified conflicting attitudes to family planning in the community, due to fear of side effects and myths about the effects of some methods, particularly intrauterine devices (IUDs). The use of condoms for STI and HIV prevention was recognized as important, but a number of barriers to correct and consistent use were identified. Violence from clients and other partners, as well as heavy alcohol use, were also highlighted.

**Figure 1.** The Women's Health Intervention using SMS for Preventing Pregnancy program logic.




### *Intervention Strategies*

The intervention was designed to incorporate specific cognitive strategies from behavior change communication theory [42,43] and appeal to women at different stages of change [34] (Table 1). Three different types of messages were used: discrete messages of less than 160 characters *pushed* to participants' phones on a predetermined schedule; role model stories, consisting of narratives about FSWs negotiating SRH risks, sent to participants' phones over several messages (episodes); and on-demand (*pull*) messages that participants could access at any time by replying to messages with specific codes.

### *Push Messages*

These messages provided specific information in less than 160 characters (1 standard SMS) and used strategies to motivate and educate participants. In the precontemplative stage, more of the messages aimed at the women were developed to be sent early in the intervention, with greater emphasis on action and maintenance later on. However, there was a mixture throughout, given the anticipated diversity of stages of change of participants. The sequencing of role model stories also reflected this approach. Push messages were delivered on alternating months to role model stories. Examples of push messages and their associated behavior change strategies are provided in Table 1.

**Table 1.** Example messages and a role model story episode mapped to behavior change theory and strategies.

| Intervention domain | Example message[a] | Stage of change and definition | Cognitive strategies |
|---|---|---|---|
| **Stand-alone push messages** | | | |
| Contraception | We have something important to tell you. Family planning lets you have sex without getting pregnant. That's what WHISPER is all about. | Precontemplation: not yet thinking about changing behavior | Increase awareness of risk; set positive outcome expectations; attract attention, *brand* recognition (social marketing strategy); frame subsequent messages |
| Dual protection | Husband or boronga (type of client)? (No matter who they are, they should be wearing a condom if they want to be with you). Hugs and kisses from WHISPER. | Precontemplation | Improve knowledge; use a friendly and personal tone to provide positive encouragement and social support; use humor to highlight desired behavior |
| Contraception | Most women who use family planning continue to have a normal sex drive. If you find one method leaves you without a sexual appetite, there are many other options. | Contemplation and preparation: thinking about making changes in behavior | Improve knowledge; challenge outcome expectations (related to fears of side effects); address specific concerns; provide an alternative strategy |
| HIV and STIs | Did you know you can take a rapid test for HIV? You get the result straight away, so you don't have to come back later! Reply 100 for services that do testing. | Contemplation and preparation | Set positive outcome expectations; motivate; provide specific action strategy |
| Gender-based violence, stigma, and rights | Violence against women is not ok, and it's not your fault. If you experience violence, remember you are not alone and can get help. Hugs, WHISPER. | Contemplation and preparation | Change social norms and model empowerment; provide social support; build self-efficacy for getting help; encourage help-seeking behavior |
| Alcohol and drug use | You can reduce your drinking: ask for beer bottles filled with water, add water to mixed drinks, secretly dump some out, drink soda, drink slow. WHISPER. | Preparation and action: preparing to act or taking actions to change behavior | Build skills and self-efficacy by breaking down behavior into components; develop action plans; encourage goal setting |
| Service information and access | If you have a bad experience with a health care provider, don't give up—ask your peer educator for clinic recommendations. Kisses and hugs. | Action and maintenance: taking actions to change behavior, for 6 months or more (maintenance) | Improve self-efficacy by overcoming setbacks; build skills to prevent or address relapse; provide alternative strategies |
| **Role model story (episode 1)** | | | |
| Contraception | Karibu tujienjoy [Welcome, let's have fun]! I'm Ciku from WHISPER. I'm new to town: I left my village because my husband drank a lot and was violent. I might be young, but I know I deserve better. I have some mpenzi [lovers] who help me out but I've had a couple of scares at the clinic, if you know what I mean. I need a better way to prevent pregnancy! | The character moving from precontemplation to contemplation in this episode. | Personalize, set scene; model self-efficacy and empowerment (leaving a violent relationship); present negative outcome expectations (risks of current behavior) |

[a]Example messages contain final content, including any modifications made during phase 2.

### *Role Model Stories*

Role modeling healthy behaviors through stories about relatable peers constitute a recognized social-cognitive strategy for behavior change [42] but have rarely been used in mHealth. The WHISPER role model stories were intended to be delivered as multiple episodes, describing FSWs who overcome barriers to contraceptive use by modeling healthy social norms and behaviors. To develop the stories, peer educators workshopped common and engaging scenarios that highlighted FSWs' risk of pregnancy (Table 2). The research team used these as the basis for developing and testing six stories, each promoting the use of LARCs integrated with other relevant themes. Peer educators also provided ideas about character names, language, and narrative, which were incorporated into the stories to ensure their relevance to the FSW community.

**Table 2.** Role model stories developed from peer educator consultation.

| Scenarios from FSW[a] peer educators | Character | Key LARC[b] method in story | Other content in the story |
|---|---|---|---|
| Moving to the city to escape a violent husband and starting in sex work | Ciku | Implant | Intimate partner violence, inconsistent protection, pregnancy, and STI scares |
| Main partner (husband or boyfriend) resisting the use of condoms and other contraception | Sandra | Implant | Part-time sex work, STI transmission from boyfriend, condom negotiation with boyfriend, contraceptive pill |
| Pressure to drink alcohol before sex with a client, and resulting adverse consequences | Lynette | IUD[c] | Sexual risk-taking while intoxicated, strategies for reducing drinking. |
| Experiencing unintended pregnancy, concern about side effects preventing the use of contraception | Olivia | IUD | Unintended pregnancy and fetal loss, rumors about different contraceptive methods |
| Being arrested and unable to access emergency contraception | Mimi | IUD | Summary of different contraceptive methods from friends and peer educator |
| Difficulty negotiating condom use with a client, and making assumptions about his STI[d] or HIV status | Joslyn | Implant | Dual method use with clients, STI myths |

[a]FSW: female sex worker.

[b]LARC: long-acting reversible contraceptive.

[c]IUD: intrauterine device.

[d]STI: sexually transmitted infection.

### *On-Demand Messages*

Previous research has indicated that messages about reproductive health that are accessible at any time via an on-demand menu are appealing and motivational to women in East Africa [33,37]. Furthermore, interventions that are bidirectional (involving both push and pull components) may have a greater impact than unidirectional messaging [26,44]. The WHISPER on-demand menu was designed to be accessed by participants sending a code via SMS at any time. They could then obtain more information about contraceptive methods and available health services in their area. On-demand messages, along with the linking and functionality of the code menu, were tested during interviews. Example on-demand messages relating to contraceptive implants are provided in Table 3.

**Table 3.** Example message providing a link to on-demand content, SMS codes, and corresponding on-demand messages.

| Push message linking to on-demand menu[a] and code to trigger message (sent via SMS) | Corresponding on-demand message[a] |
|---|---|
| **WHISPER Alert: Mimi found out that the implant protects you from pregnancy for 3 to 5 years! Want to know more about the implant? Reply 11 to this text** | |
| 11 | Implants are small rods placed under the skin of a woman's arm. Highly effective for 3 to 5 years. It can be removed anytime. For married and singles. May cause light irregular bleeding. When removed, you can become pregnant with no delay. No infertility or birth defects. WHISPER main menu, reply 00. For more information, reply 12. |
| 12 | Implants: Benefits 13, Side effects 14, Bleeding side effects 15, True facts 16, Insertion/Removal 17, WHISPER main menu 00. |
| 15 | Implant bleeding side effects: Monthly bleeding may be lighter and irregular during the first year, then lighter, more regular, and infrequent. Monthly bleeding may stop for some women. If bleeding stops it is not harmful—blood does not build up inside the womb. |

[a]Example messages contain final content, including any modifications made during phase 2.

## Phase 2: Testing and Refining Messages

The second phase involved testing and refining draft messages based on detailed feedback from FSWs in 6 workshops and 12 interviews. A semistructured workshop guide covered responses to specific messages, overall feedback on the program, and preferences for program structure, timing, delivery, and vernacular. A female researcher who had experience working with FSWs was employed to moderate the workshops rather than an FSW peer educator because this provided more anonymity and privacy to participants. Also, experienced facilitating groups was considered necessary to cover content efficiently but in a sensitive way. A note-taker was also present.

Workshop participants were recruited through purposive sampling of sex work venues (*hotspots*) by FSW peer educators. Peer educators with connections to different hotspots across two subcounties of Mombasa (Kisauni and Changamwe) were selected to be recruiters. Recruiters sought participants with a range of ages and education levels. To be eligible for the

workshops, women had to be at least 16 years old, have received money or goods in exchange for sex in the previous 6 months, self-report that they were not currently pregnant or planning a pregnancy in the next year, and own and use a mobile phone. These criteria were consistent with those in the subsequent RCT [29].

We modified the intervention based on the results of the workshops and tested it in 12 one-on-one interviews with participants who met the same eligibility criteria. Interviews tested the technical performance of the SMS system, reactions to the use of on-demand (*pull*) messages, and interpretation of specific messages where there was uncertainty about meaning.

Workshops and interviews were audio-recorded, and detailed notes taken during the sessions were augmented with data from the recordings and translated into English, where necessary, by research assistants. A mixed deductive and inductive thematic approach to analysis was adopted [45]. A list of predetermined codes was used to obtain specific information about message delivery, wording, and preferred content domains. Other codes emerged from the data and were analyzed thematically, particularly in relation to how FSWs responded to the content of different messages and related this to their own experiences, and how behavior change strategies employed by the messages resonated with participants. We analyzed the data using NVivo 11 (QSR International Pty Ltd).

Participants of both workshops and interviews provided written consent before proceeding. They were provided with refreshments and given 500 Kenyan shillings (approximately US $5) to reimburse them for their time and travel costs. The study was approved by the Monash University Human Research Ethics Committee (Australia) and the University of Nairobi and Kenyatta Hospital Ethics Committee (Kenya).

Following data analysis, we refined the intervention further by making recommended wording changes, emphasizing certain content, finalizing the structure (order, timing, and frequency of messages), and resolving technical implementation issues.

# Results

## Workshop Participant Characteristics

We held 6 workshops, each with 7 FSWs, in November 2015 to test the draft messages and refine the intervention. Workshops A and E were held solely with women who had experienced unintended pregnancy to allow open discussion of this issue. Most participants were in their mid-20s (median age 24 years, IQR 20-30) with some secondary education (secondary: 20/42, 48%; primary 15/42, 36%; tertiary: 7/42, 17%) and at least 1 child (34/42, 81%). They worked from a range of hotspots, with half working from bars or nightclubs. Just over half of participants owned a smartphone (the remainder had feature or basic mobile phones), and almost all used SMS at least daily. It was common for participants to share text messages (35/42, 83%), mostly with friends, and some also with family, boyfriends, and clients.

## Responses to Message Content

### *Importance of Topics and Relevance*

Participants felt that the topics covered were high priorities for FSWs and would be useful to their community. They confirmed that unintended pregnancy was an important issue that caused fear and stress, and relayed personal experiences of getting pregnant unintentionally. They particularly liked messages that gave general pregnancy prevention advice and information about IUDs.

> *Many sex workers fear getting pregnant [more] than HIV.* [Age 19, Kisauni, workshop F]
>
> *Message 2 is important…as female sex workers we must use family planning because we have many clients and we need to protect ourselves from becoming pregnant.* [Age 20, Kisauni, workshop E]

There was a strong positive response to messages on rights, violence, and alcohol use, particularly when violence hotlines were provided, and practical tips were given to reduce alcohol-related harms:

> *Many sex workers do not know their rights so by sharing with them [these] messages they will be informed.* [Age 24, Changamwe, workshop D]

Messages were considered highly relevant and spoke to participants' real experiences, particularly the role model stories, with which participants strongly identified.

> *This information talks about what sex workers go through.* [Age 19, Changamwe, workshop F]

Many women volunteered personal stories that echoed message content. Common scenarios were pregnancy scares, difficulty negotiating condom use, experiences of violence, and contraceptive side effects. Role model stories in which the character gets drunk and then needs to use emergency contraception, and in which a woman overcomes contraceptive myths to use an IUD prompted the most discussion and personal anecdotes.

> *It is realistic. I had the experience when the condom busted and I was unable to access e-pills on time, therefore I conceived a baby and I had no option of aborting, therefore I carried pregnancy to term.* [Age 35, Kisauni, workshop E]
>
> *This thing happens to sex workers and it has happened to me, too.* [Referring to unprotected sex while drunk; age 36, Kisauni, workshop C]

### *Appeal and Tone*

Most women found the messages interesting and appealing, and several commented that the messages stimulated an interest in them to find out more. The majority in all groups agreed that they felt inspired by the role model stories.

> *It is inspirational especially when Sandra [character in a story] visits a health center for screening and also consults friends on STI prevention.* [Age 40, Changamwe, workshop B]

> *I am inspired. It shows us different family planning methods for example depo, IUD.* [Age unknown, Kisauni, workshop E]

Positive tone also contributed to the appeal. When specifically asked about tone, the most common responses were that the messages were friendly (mentioned 21 times), educational (16), and polite (9). Six women also commented spontaneously that the messages were caring:

> *They are friendly because they let you know that there is someone who cares for you.* [Age 44, Changamwe, workshop B]

> *It is friendly. The message is like peers talking to me. It is not official.* [Age 35, Kisauni, workshop E]

Participants were asked if FSWs would trust the information provided. Most agreed they would, because the messages were caring and relevant to sex workers, and their community had been involved in developing them.

> *They will trust [the information] because somebody is caring for them.* [Age 22, Changamwe, workshop F]

> *This information is good and they will accept it and also [because] we have been involved.* [Age 24, Changamwe, workshop B]

## Responses to Behavior Change Strategies

Behavior change strategies adopted from social cognitive theory that were used to develop messages resonated with women. Strategies that were most strongly echoed in their responses were the provision of knowledge, change in outcome expectations, self-efficacy and skill development, and empowerment.

### *Knowledge Gain*

A large number of participants reflected that the messages taught them new and useful information. This was the response to both the program overall and specific topics, particularly messages on contraceptive options and side effects, IUDs, condoms, HIV, and alcohol. Participants from workshop A, who were less educated than other groups, were particularly keen to learn more.

> *I would like to learn more so I would enroll [in the program].* [Age 23, Changamwe, workshop A]

> *Friends would want to know more…Yes I will be taught then share the information, especially among sex workers on unwanted pregnancies.* [Age 22, Kisauni, workshop D]

Specific knowledge gaps were identified as negatively affecting individual participants and their community. Knowledge gaps in HIV transmission were mentioned 4 times, condom usage techniques 3 times, side effects of family planning twice, appropriateness of using IUD with multiple partners twice, menstrual cycles twice, and alcohol and rights once each.

Women frequently mentioned how messages challenged prevalent myths about contraception, particularly about side effects and appropriate use of IUDs:

> *I did not know that one can use a coil [IUD] and still have many partners.* [Age 23, Changamwe, workshop A]

> *I can relate to this episode because I knew with sperms my sitting allowance [buttocks] would increase and my side mirror [hips] would expand, but that was a myth. I have learnt.* [Referring to myth that sperm in the vagina is beneficial; age 19, Kisauni, workshop F]

However, some described incorrect ideas that they or their peers still held about contraception:

> *The coil is not good for sex workers because of the nature of work. We have different men of different [penis] sizes.* [Age 38, Kisauni, workshop E]

Only 2 participants stated that they did not learn anything new, indicating that the level of information was generally well targeted to participants' background knowledge.

### *Outcome Expectations*

*Outcome expectations* refers to the beliefs one holds about the outcomes that will result from a specific behavior [42]. Many messages triggered participants to think about the outcomes of their behaviors, both positive and negative, particularly in relation to family planning. They were also prompted to think about the outcomes of heavy alcohol use, STI prevention, and service utilization. Examples of the outcomes they reflected on are presented in Table 4.

**Table 4.** Outcome expectations raised by workshop participants and corresponding quotes.

| Outcome expectations | Example quotes |
|---|---|
| An outcome of using family planning is not getting pregnant and hence avoiding related stress | • "If I am with a client and I am on family planning, I will not fear issues of pregnancy." (Age 24, Changamwe, workshop B)<br>• "When she [character in a role model story] uses coil, she is free and does not have fear of getting pregnant." (Age 23, Kisauni, workshop D) |
| Some contraceptive methods cause negative outcomes in the form of side effects but these are less severe than many perceive and can be addressed | • "These contraceptives have different effects: they lie to the body that you are pregnant [due to amenorrhea], but if you know the effects there is need not to worry." (Age 20, Kisauni, workshop D)<br>• "It has inspired me, because I can use a coil and when I want a baby I can return to fertility and conceive." (Age 19, Kisauni, workshop E) |
| Getting drunk results in increased risk and bad business | • "If I get drunk when I go to the hotspot I will not be able to negotiate well with the client and I might be violated. I will not be able to get what I wanted." (Age 19, Changamwe, workshop F)<br>• "When I am sober I will take care of myself from drama and keep myself safe, as sometimes men take advantage if one is drunk; he may refuse to pay you, steal your money and phone, or even not use a condom." (Age 36, Kisauni, workshop C) |
| If one accesses a service, they can expect to be provided with good quality care | • "When I go the clinic I can get help for an implant or STI treatment." (Age 20, Changamwe, workshop A)<br>• "I have learnt that a health worker can listen to a sex worker and give advice." (Age 24, Changamwe, workshop D) |

Role model stories appeared particularly well suited to supporting changes in outcome expectations and triggered responses in which women reflected on the behaviors of the characters, and the outcomes of their own behaviors and those of peers:

> *Yes; Lynette [character in story] was drunk and did not have a family planning method, if she had coil the situation could have been avoided.* [Age 30, Kisauni, workshop C]

> *My friend had a fear of using a coil, but when she went to hospital she was given advice and more information and she ended up using it, and it is not disturbing her.* [Age 30, Kisauni, workshop E]

### *Self-Efficacy, Skills, and Action*

Participants' comments indicated a belief that they or their peers are capable of adopting certain behaviors, demonstrating self-efficacy for healthy behavior. They also reflected that some of the messages improved their skills and confidence to adopt new behaviors. Messages that provided specific skills and techniques to lower drinking risk, and specific tips on condom use, were particularly well-received:

> *I can talk to the waiter and exchange beer with water.* [Age 26, Kisauni, workshop C]

Violence and rights messages prompted statements reflecting increased self-efficacy for recognizing rights, seeking help, and negotiating with clients:

> *I know now I am the boss and I can negotiate for payment with clients.* [Age 22, Kisauni, workshop F]

> *These messages will teach them their rights, and how they can negotiate and report cases if violated.* [Age 18, Kisauni, workshop F]

Messages on what to expect from service providers also prompted a response that suggested women felt capable of accessing services:

> *It is true—the job of the health care is to give services, and I can find a clinic where I am comfortable.* [Age 26, Kisauni, workshop C]

Participants liked messages that suggested specific actions or plans, and they were triggered to think about what they should do in different situations and how they could make the best use of the messages:

> *I will put a reminder on the message that I have received, for example when I am at the hotspot.* [Age 19, Kisauni, workshop F]

> *If one contraceptive is not good for you, change to another one.* [Age 22, Kisauni, workshop D]

> *When I go out I should have a friend or talk to the receptionist at the hotspot to check on my security, and not go with the money in the room.* [Age 30, Kisauni, workshop C]

### *Empowerment*

Empowerment refers to a process in which individuals and communities gain control over their lives and the issues that most affect them, and includes the development of self-confidence and self-reliance [46-48]. The responses of women indicate that they found the messages empowering to both themselves and their community, particularly messages about violence and rights:

> *It is about me, myself and I. I deserve to be happy and know my rights. Yes I like this message [about rights of sex workers].* [Age 24, Changamwe, workshop D]

> *We should visit people who can listen to our voice or our complaints, and health workers should not stigmatize us when we go for services.* [Age 23, Kisauni, workshop D]

There was a sense that the messages prompted improved morale and inspired them to take action. A number of women

specifically mentioned the importance of being in control, particularly in response to role model stories. Stories about the use of LARCs also prompted a sense of being free from the fear of getting pregnant:

> *They give me morale to use condoms.* [Age 31, Changamwe, workshop A]
>
> *Dual methods remove fear. I have total control when I have the implant and use condoms.* [Age 23, Changamwe, workshop A]

Participants were overwhelmingly in favor of sharing the messages with other sex workers and friends, and to a lesser extent, with family members, boyfriends, clients, and health workers. Almost all said they would share messages when asked directly, and many said that they would do so without prompting, consistent with the existing practice of frequent sharing. The desire to share influenced their preferences for message delivery. Participants in workshop E preferred SMS because it is an easy format to share. Those in workshops A and B wanted to receive the messages before starting work, to allow time to discuss them with others at the hotspot. Many indicated that it was important for both sex workers and the broader community to have access to this information, and that, as holders of the messages, they would be empowered to provide it. There was a real enthusiasm expressed for teaching others:

> *My friends do not have this information, therefore I will reach out to them and share with them.* [Age 31, Changamwe, workshop A]
>
> *I will share with 15 and 16 year age groups, because they do not know about family planning and they are already engaging in sex.* [Age 20, Kisauni, workshop E]
>
> *I will share with my clients so that they can reach their spouses.* [Age 33, Kisauni, workshop E]
>
> *By teaching others these messages they will help me to remember.* [Age 16, Changamwe, workshop F]

### Risks of the Program

One workshop participant thought that she could contact WHISPER to receive emergency assistance ("If am assaulted I can send message or call to get help"; age 40, Kisauni, workshop C). As WHISPER is an automated system, such requests cannot be followed up, and it was concerning that the women may have thought they could depend on the program in this way. This was addressed in subsequent changes to the program (described below).

Breach of privacy was also raised as a potential risk. Participants in 3 workshops discussed the risk that someone else would see the messages and would assume that they were sex workers and/or HIV positive. Some were afraid that this could cause conflict with their boyfriends.

> *It will bring conflict between me and my boyfriend who might be nosy especially on information on STIs.* [Age 24, Changamwe, workshop B]
>
> *It depends on the person and the relationship you have with them. For example, if a parent sees information about a condom he or she will react, but you can explain. If a client or a boyfriend sees information on HIV he will panic.* [Age 19, Kisauni, workshop F]

However, not all agreed, and there was a discussion about how the messages might be good for other people, including their boyfriends, illustrated in this interaction:

> *Even the boyfriends want to plan a family, so they cannot deter us from using this service.* [Age 35, Changamwe, workshop B]
>
> *These messages will be good for both parties—man and woman.* [Age 40, Changamwe, workshop B]

Others felt that the messages would be socially acceptable. For example, workshop A participants thought friends and health care workers would be impressed that they were *careful with their lives*:

> *My boyfriend, family or friend will say I am informed.* [Age 23, Changamwe, workshop A]

Another risk is that the program would not overcome barriers to healthy behaviors in sex work. Responses illustrated how some barriers cannot be overcome by an individually targeted intervention alone. For example, a role model story about a client offering to pay extra for no condom prompted discussion in workshop F about the need to balance conflicting outcome expectations of different courses of action. This reflected sex workers' need to continually assess risk, and the fact that money and immediate safety are often higher priorities than pregnancy and STI prevention.

> *The client of Joslyn [character in the story] in this case was polite, because he said he will call next week, but most clients will become abusive if you refuse to not use condoms.* [Age 19, Changamwe, workshop F]
>
> *The issue is money. That is why female sex workers risk going without a condom—so that she might get a client.* [Age 19, Kisauni, workshop F]
>
> *I had a friend who had the same issue. She judged the guy with looks because the guy had money. She did not negotiate for condom before. The money was huge. The lady refused because this guy insisted no condom.* [Age 22, Kisauni, workshop F]

### Response to On-Demand Messages

Interview participants were sent messages with a link to the on-demand system. In all, 7 of 12 participants found it *very easy* to access messages on demand. Others had minor difficulties, and 2 had genuine difficulty and had to be directed by the interviewer. These women had lower education than other participants.

Many women liked having the option to retrieve more information and the interactive aspect of the system. They talked about the ease of getting detailed information on their phones rather than having to seek it out from health professionals, and the ability to refer back to such messages later. A number of women did not feel the initial message on a topic contained new information, but obtaining more detail allowed them to gain a greater understanding.

> *It has motivated me since I can get instant replies and can be helped instantly.* [Age 26, Changamwe]
>
> *It is like revision [on] family planning when I am reading I am being enlightened more and remembering, it is easy.* [Age 33, Changamwe]

Messages about health services were considered very useful to participants and their peers as they provided information that was not easy to obtain and saved the time and resources needed to find appropriate services. The emergency message (developed in response to workshop feedback—described below) was particularly popular and seen as important.

> *When I need help, or having an emergency, they have provided a number which I can call for free in case of violence and has given me a whisper menu too. In short they have not left me hanging from the situation I may be experiencing.* [Age 30, Kisauni]

There were some technical issues during interviews, including delayed receipt of messages and (erroneous) warnings received from network providers, which deterred some women from continuing. Despite these challenges, most interview participants were very engaged in the process of retrieving pull messages, and those who had initial difficulty still enjoyed the process. When asked directly, all agreed that they would like to continue using the system.

### Intervention Structure and Final Delivery

#### *Intervention Delivery Preferences*

Workshop participants had generally consistent preferences regarding how the intervention should be delivered. The majority were in favor of text rather than voice modality and preferred push messages to retrieving content via a pull system. Most women reported that their texting practices involved a mixture of English and Swahili, and they favored English for health messages, with some keywords or phrases in Swahili. Participants wanted to receive messages several times a week for at least 1 year and preferred to receive them in the late morning on set days to align with their typical work schedules.

#### *Refinement of the Intervention*

A number of changes were made to the intervention content and form based on findings. To minimize the risk of women expecting emergency assistance from WHISPER, a message was included on what to do in an emergency, specifically around violence. An *error* message was also developed that was triggered if they tried to send content other than the prespecified codes. These were well received on testing in the interviews.

The other key concern identified was the risk of sex work status being discovered by clients or boyfriends viewing the messages. In response to this, we minimized overt references to sex work and clients wherever possible.

Suggestions were adopted from participants regarding the use of specific words and terms, in both Swahili and English. Terms of endearment like *mrembo* (beautiful) and *darling* were incorporated into the messages, and *family planning* was adopted consistently as FSW's preferred term for pregnancy prevention.

A number of strategies were adopted to address the technical challenges encountered using the on-demand system. These included testing the system with each participant during their enrolment and incorporating introductory messages that explained how to use the on-demand menu.

#### *Final Intervention Schedule*

The intervention components and delivery schedule were finalized based on workshop and interview results. Over a 12-month intervention period, participants received SMS 2 to 3 times per week, alternating push messages with role model stories every month. A total of 82 push messages were developed for the intervention (see examples in Table 1). In addition, 7 reminders for study visits and 19 alerts linking to the pull system (Table 2) were sent to participants. Six role model stories were sent, each with 4 or 5 episodes (Table 1). Messages were scheduled for mornings on set days, in line with participant preferences.

## Discussion

We provide the first description of the development of a digital health intervention for FSWs that uses a comprehensive SRHR framework. The participatory approach enabled FSWs to influence the range and content of topics included in the intervention [49] and to enhance the relevance and salience of messages, and the participants themselves confirmed that their involvement improved the perceived trustworthiness of the messages. The benefits of a co-design approach have been observed in other mHealth studies [32]. Co-design is critical for handling sensitive content matter that may be interpreted differently by different communities [50].

Furthermore, health behavior change interventions are more effective when they are based on social and behavioral science theory, and the use of multiple theories may increase intervention effectiveness [51,52]. WHISPER utilizes multiple theories to guide the intervention framework and specify intermediate behavior change outcomes [42,43]. Notably, the adopted strategies from Bandura's social cognitive theory [35] were frequently highlighted by workshop participants, confirming the applicability and utility of theory for guiding intervention design.

The messages increased participants' feelings of empowerment [46-48] and social support [43]. There was a strong sense of being part of a community; many women reflected on how messages would help their friends, or how they could share the knowledge they had gained, rather than focusing solely on how it would help them as individuals. These findings suggest that WHISPER may capitalize on and enhance community cohesion. Social cohesion has been linked to safer sex behaviors [53] and is important for the success of community empowerment interventions, which may otherwise be undermined by mistrust and competition among FSWs for scarce resources [7]. The desire to share messages with peers and the broader community suggests that social diffusion is also likely to contribute to the effectiveness of the intervention [43].

Participants reinforced the importance of the selected SRHR topics and confirmed that unintended pregnancy is a major

concern for sex workers. The team was careful to ensure that scenarios were not overly optimistic and appropriately represented known barriers. Content addressing family planning myths was stated in different ways and different formats (push and pull messages and role model stories) to maximize the potential that participants would engage with and learn from the WHISPER content so that myths would no longer represent barriers to participants.

In addition to family planning, alcohol use and gender-based violence were viewed as important. Strategies for reducing drinking provided in the text messages were adapted from effective harm reduction interventions [54], and this practical emphasis resonated strongly with participants. Experiences of violence were frequently described and noted as a barrier to adopting safer sexual practices. Although an individual health promotion intervention cannot address the structural causes of violence or change the behavior of perpetrators, participant responses indicate that messages about violence improved knowledge of rights, were empowering, and provided much-needed advice about how to reduce risks and access services.

The intervention was highly acceptable to both workshop and interview participants. Women were interested and engaged in both the content and the format of delivery, with role model stories eliciting particularly enthusiastic discussion, and SMS confirmed as the preferred technology. Workshop and interview participants demonstrated familiarity and comfort using SMS, and desire to learn more, suggesting that it is feasible for SRHR messages to be sent regularly over a year to this population. Testing during interviews confirmed the feasibility of the on-demand system. Most participants could retrieve pull messages with relative ease; however, women who are less educated or have less experience with mobile phones may experience difficulty using this system.

There were some technical issues, including a network warning that could not be deactivated. Similar problems have been identified by other implementers of mHealth programs [55,56], highlighting the importance of real-time testing and the need to consider how to overcome aspects of mobile platforms designed for commercial rather than public health applications.

We have demonstrated that WHISPER is feasible to implement and acceptable to the target audience; however, this may not translate to sufficient participant engagement to produce better health outcomes. Engagement with a digital health program incorporates not only the subjective and cognitive responses that are triggered (which are explored in this paper) but also the extent of use [57] (eg, the number and frequency of messages received), which will be measured during the trial. The evaluation of engagement in digital health interventions has not been well characterized and is an important area for further research [57].

Our research revealed several risks to participation in a digital health SRHR intervention. First, participants believed that they could receive emergency assistance from WHISPER. It is possible that the friendly and personal tone—while effective in generating intervention engagement [58]—creates an expectation that participants are interacting with real people rather than an automated system. Revisions were made to minimize this risk.

Second, disclosure of sensitive messages could result in increased conflict with boyfriends or clients, although it also has the potential to improve communication with partners. Disclosure risk has been explored during the development of mHealth interventions for HIV [58-60], and an increase in intimate partner violence was an unintended consequence of a contraceptive mHealth program in Bangladesh [61]. However, few studies report on the potential harms from women's participation in SRHR digital health programs, and this is an important area for further research [61,62].

This study had some important limitations. We used purposive sampling and cannot ensure that workshop participants were representative of the larger FSW population. In addition, our approach to data collection and analysis was highly directive, and some messages were not tested because they were from preexisting mHealth interventions [37]. This approach to data collection yielded the specific information needed to develop the intervention, but it was not designed to reach data saturation, and it is possible that some critical feedback was not obtained.

This research provides a clear illustration of the many issues that preoccupy FSWs in their day-to-day lives—beyond the traditional biomedical focus on HIV risk and transmission. Our results support the need for health promotion interventions that utilize a participatory approach to intervention development and are based on social and behavioral science to increase their relevance and effectiveness. The resulting WHISPER digital health intervention reflects the complex reality of FSWs' daily lives and provides a feasible, engaging, and confidential approach for addressing their SRHR concerns and needs.

## Acknowledgments

This work was supported by the National Health and Medical Research Council Australia (NHMRC) under Project Grant GNT 1087006, NHMRC Career Development Fellowships for SL and MS, and an NHMRC Postgraduate Scholarship for FA. The authors would like to acknowledge the FSW peer educators and research assistants who contributed to this work. We also gratefully acknowledge the contribution to this work of funding from the Victorian Operational Infrastructure Support Program received by the Burnet Institute.

## Conflicts of Interest

None declared.

## References


1. Dhana A, Luchters S, Moore L, Lafort Y, Roy A, Scorgie F, et al. Systematic review of facility-based sexual and reproductive health services for female sex workers in Africa. Global Health 2014 Jun 10;10:46 [FREE Full text] [doi: 10.1186/1744-8603-10-46] [Medline: 24916010]
2. National AIDS/STD Control Programme of the Kenya Ministry of Public Health and Sanitation. Bar Hostess Empowerment & Support Programme (BHESP). 2010 Sep. Standards for Peer-Education and Outreach Programs for Sex Workers URL: https://tinyurl.com/y7j5kjnx [accessed 2020-03-10]
3. Luchters S, Chersich MF, Rinyiru A, Barasa M, King'ola N, Mandaliya K, et al. Impact of five years of peer-mediated interventions on sexual behavior and sexually transmitted infections among female sex workers in Mombasa, Kenya. BMC Public Health 2008 Apr 29;8:143 [FREE Full text] [doi: 10.1186/1471-2458-8-143] [Medline: 18445258]
4. Chersich MF, Luchters S, Ntaganira I, Gerbase A, Lo YR, Scorgie F, et al. Priority interventions to reduce HIV transmission in sex work settings in sub-Saharan Africa and delivery of these services. J Int AIDS Soc 2013 Mar 4;16:17980 [FREE Full text] [doi: 10.7448/IAS.16.1.17980] [Medline: 23462140]
5. Lafort Y, Geelhoed D, Cumba L, Lázaro CD, Delva W, Luchters S, et al. Reproductive health services for populations at high risk of HIV: Performance of a night clinic in Tete province, Mozambique. BMC Health Serv Res 2010 May 28;10:144 [FREE Full text] [doi: 10.1186/1472-6963-10-144] [Medline: 20507644]
6. Yam EA, Okal J, Musyoki H, Muraguri N, Tun W, Sheehy M, et al. Kenyan female sex workers' use of female-controlled nonbarrier modern contraception: do they use condoms less consistently? Contraception 2016 Mar;93(3):222-225. [doi: 10.1016/j.contraception.2015.11.010] [Medline: 26656841]
7. Cornish F, Campbell C. The social conditions for successful peer education: a comparison of two HIV prevention programs run by sex workers in India and South Africa. Am J Community Psychol 2009 Sep;44(1-2):123-135. [doi: 10.1007/s10464-009-9254-8] [Medline: 19521765]
8. Petruney T, Minichiello SN, McDowell M, Wilcher R. Meeting the contraceptive needs of key populations affected by HIV in Asia: An unfinished agenda. AIDS Res Treat 2012;2012:792649 [FREE Full text] [doi: 10.1155/2012/792649] [Medline: 22991656]
9. Moore L, Chersich MF, Steen R, Reza-Paul S, Dhana A, Vuylsteke B, et al. Community empowerment and involvement of female sex workers in targeted sexual and reproductive health interventions in Africa: a systematic review. Global Health 2014 Jun 10;10:47 [FREE Full text] [doi: 10.1186/1744-8603-10-47] [Medline: 24916108]
10. Slabbert M, Venter F, Gay C, Roelofsen C, Lalla-Edward S, Rees H. Sexual and reproductive health outcomes among female sex workers in Johannesburg and Pretoria, South Africa: Recommendations for public health programmes. BMC Public Health 2017 Jul 4;17(Suppl 3):442 [FREE Full text] [doi: 10.1186/s12889-017-4346-0] [Medline: 28832290]
11. Beckham SW, Shembilu CR, Brahmbhatt H, Winch PJ, Beyrer C, Kerrigan DL. Female sex workers' experiences with intended pregnancy and antenatal care services in southern Tanzania. Stud Fam Plann 2015 Mar;46(1):55-71 [FREE Full text] [doi: 10.1111/j.1728-4465.2015.00015.x] [Medline: 25753059]
12. Global Health Council. Washington, DC: The Center for Health and Gender Equity; 2016 Aug 5. REPORT: All Women, All Rights, Sex Workers Included: US Foreign Assistance and the Sexual and Reproductive Health and Rights of Female Sex Workers URL: https://tinyurl.com/y8zk6tft [accessed 2020-03-10]
13. Ampt FH, Willenberg L, Agius PA, Chersich M, Luchters S, Lim MS. Incidence of unintended pregnancy among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. BMJ Open 2018 Sep 17;8(9):e021779 [FREE Full text] [doi: 10.1136/bmjopen-2018-021779] [Medline: 30224388]
14. Sutherland EG, Alaii J, Tsui S, Luchters S, Okal J, King'ola N, et al. Contraceptive needs of female sex workers in Kenya - a cross-sectional study. Eur J Contracept Reprod Health Care 2011 Jun;16(3):173-182. [doi: 10.3109/13625187.2011.564683] [Medline: 21413869]
15. Cheuk E, Becker M, Mishra S, Blanchard J, Isac S, Kingola N, et al. Understanding Female Sex Workers' Early HIV Risk and the Implications for HIV Epidemic Control (Transitions Study): Mapping and Estimating the Population Size of Female Sex Workers in Mombasa, Kenya and Dnipropetrovsk, Ukraine. In: Proceedings of the 24th Annual Canadian Conference on HIV/AIDS Research. 2015 Presented at: CAHR'15; April 30 - May 3, 2015; Toronto, Ontario URL: https://www.cahr-acrv.ca/wp-content/uploads/2012/10/InfDis_26_SB_MarApr2015_Final.pdf
16. Luchters S, Bosire W, Feng A, Richter ML, King'ola N, Ampt F, et al. 'A baby was an added burden': predictors and consequences of unintended pregnancies for female sex workers in Mombasa, Kenya: a mixed-methods study. PLoS One 2016;11(9):e0162871 [FREE Full text] [doi: 10.1371/journal.pone.0162871] [Medline: 27689699]
17. Khan MR, Turner AN, Pettifor A, van Damme K, Rabenja NL, Ravelomanana N, Mad STI Prevention Group, et al. Unmet need for contraception among sex workers in Madagascar. Contraception 2009 Mar;79(3):221-227 [FREE Full text] [doi: 10.1016/j.contraception.2008.09.011] [Medline: 19185677]
18. Boudreau CL. Assessing the contraceptive needs of female sex workers in Kigali, Rwanda. Ann Glob Health 2015;81(1):217. [doi: 10.1016/j.aogh.2015.02.997]
19. Okal J, Stadler J, Ombidi W, Jao I, Luchters S, Temmerman M, et al. Secrecy, disclosure and accidental discovery: perspectives of diaphragm users in Mombasa, Kenya. Cult Health Sex 2008 Jan;10(1):13-26. [doi: 10.1080/13691050701519730] [Medline: 18038278]

20. Williamson LM, Parkes A, Wight D, Petticrew M, Hart GJ. Limits to modern contraceptive use among young women in developing countries: a systematic review of qualitative research. Reprod Health 2009 Feb 19;6:3 [FREE Full text] [doi: 10.1186/1742-4755-6-3] [Medline: 19228420]
21. Sedgh G, Hussain R, Bankole A, Singh S. Guttmacher Institute. New York: Guttmacher Institute; 2007 Jun. Women With an Unmet Need for Contraception in Developing Countries and Their Reasons for Not Using a Method URL: https://www.guttmacher.org/sites/default/files/pdfs/pubs/2007/07/09/or37.pdf [accessed 2020-03-10]
22. Thomas B, Closson EF, Biello K, Menon S, Navakodi P, Dhanalakshmi A, et al. Development and open pilot trial of an HIV-prevention intervention integrating mobile-phone technology for male sex workers in Chennai, India. Arch Sex Behav 2017 May;46(4):1035-1046 [FREE Full text] [doi: 10.1007/s10508-015-0665-3] [Medline: 26714684]
23. Information and Communication Technology (ICT). 2018. Statistics URL: http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx [accessed 2019-07-09]
24. Smith C, Gold J, Ngo TD, Sumpter C, Free C. Mobile phone-based interventions for improving contraception use. Cochrane Database Syst Rev 2015 Jun 26(6):CD011159 [FREE Full text] [doi: 10.1002/14651858.CD011159.pub2] [Medline: 26115146]
25. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
26. Rokicki S, Cohen J, Salomon JA, Fink G. Impact of a text-messaging program on adolescent reproductive health: a cluster-randomized trial in Ghana. Am J Public Health 2017 Feb;107(2):298-305. [doi: 10.2105/AJPH.2016.303562] [Medline: 27997236]
27. Harrington EK, Drake AL, Matemo D, Ronen K, Osoti AO, John-Stewart G, et al. An mHealth SMS intervention on postpartum contraceptive use among women and couples in Kenya: a randomized controlled trial. Am J Public Health 2019 Jun;109(6):934-941. [doi: 10.2105/AJPH.2019.305051] [Medline: 31067089]
28. Johnson D, Juras R, Riley P, Chatterji M, Sloane P, Choi SK, et al. A randomized controlled trial of the impact of a family planning mHealth service on knowledge and use of contraception. Contraception 2017 Jan;95(1):90-97. [doi: 10.1016/j.contraception.2016.07.009] [Medline: 27421767]
29. Ampt FH, Mudogo C, Gichangi P, Lim MS, Manguro G, Chersich M, et al. WHISPER or SHOUT study: protocol of a cluster-randomised controlled trial assessing mHealth sexual reproductive health and nutrition interventions among female sex workers in Mombasa, Kenya. BMJ Open 2017 Aug 18;7(8):e017388 [FREE Full text] [doi: 10.1136/bmjopen-2017-017388] [Medline: 28821530]
30. O'Mara-Eves A, Brunton G, McDaid D, Oliver S, Kavanagh J, Jamal F, et al. Community engagement to reduce inequalities in health: a systematic review, meta-analysis and economic analysis. Pub Health Res 2013;1(4) [FREE Full text] [Medline: 25642563]
31. Cyril S, Smith BJ, Possamai-Inesedy A, Renzaho AM. Exploring the role of community engagement in improving the health of disadvantaged populations: a systematic review. Glob Health Action 2015;8:29842 [FREE Full text] [doi: 10.3402/gha.v8.29842] [Medline: 26689460]
32. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
33. L'Engle K, Plourde KF, Zan T. Evidence-based adaptation and scale-up of a mobile phone health information service. Mhealth 2017;3:11 [FREE Full text] [doi: 10.21037/mhealth.2017.02.06] [Medline: 28567408]
34. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983 Jun;51(3):390-395. [doi: 10.1037//0022-006x.51.3.390] [Medline: 6863699]
35. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Englewood Cliffs, NJ: Prentice Hall; 1986.
36. Lim MS, Hocking JS, Aitken CK, Fairley CK, Jordan L, Lewis JA, et al. Impact of text and email messaging on the sexual health of young people: a randomised controlled trial. J Epidemiol Community Health 2012 Jan;66(1):69-74. [doi: 10.1136/jech.2009.100396] [Medline: 21415232]
37. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
38. Vahdat HL, L'Engle KL, Plourde KF, Magaria L, Olawo A. There are some questions you may not ask in a clinic: providing contraception information to young people in Kenya using SMS. Int J Gynaecol Obstet 2013 Nov;123(Suppl 1):e2-e6 [FREE Full text] [doi: 10.1016/j.ijgo.2013.07.009] [Medline: 24012514]
39. World Health Organization. Family Planning - A Global Handbook for Providers. Baltimore and Geneva: CCP and WHO; 2011.
40. Ministry of Public Health and Sanitation. National Family Planning Guidelines for Service Providers: Updated to reflect the 2009 Medical Eligibilty Criteria of the World Health Organization. Nairobi: Division of Reproductive Health, Ministry of Public Health and Sanitation; 2010.

41. Ministry of Public Health Sanitation. National guidelines for HIV/STI programs for sex workers. Nairobi, Kenya: Ministry of Public Health Sanitation; 2010.
42. Maibach E, Cotton D. Moving people to behaviour change. In: Maibach EW, Parrott RL, editors. Designing Health Messages: Approaches from Communication Theory and Public Health Practice. Thousand Oaks: Sage publications; 1995.
43. National Cancer Institute. Division of Cancer Control and Population Sciences. Bethesda, MD: National Institutes of Health; 2005. Theory at a Glance: A Guide for Health Promotion Practice URL: https://cancercontrol.cancer.gov/brp/research/theories_project/theory.pdf [accessed 2020-03-10]
44. Armanasco AA, Miller YD, Fjeldsoe BS, Marshall AL. Preventive health behavior change text message interventions: a meta-analysis. Am J Prev Med 2017 Mar;52(3):391-402. [doi: 10.1016/j.amepre.2016.10.042] [Medline: 28073656]
45. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
46. Wallerstein N. Powerlessness, empowerment, and health: implications for health promotion programs. Am J Health Promot 1992;6(3):197-205. [doi: 10.4278/0890-1171-6.3.197] [Medline: 10146784]
47. Canadian International Development Agency. Cida's Policy On Gender Equality. Quebec: CIDA; 1999.
48. Varkey P, Mbbs, Kureshi S, Lesnick T. Empowerment of women and its association with the health of the community. J Womens Health (Larchmt) 2010 Jan;19(1):71-76. [doi: 10.1089/jwh.2009.1444] [Medline: 20088661]
49. World Health Organization. Community Participation in Local Health and Sustainable Development: Approaches and Techniques. Copenhagen: World Health Organization Regional Office for Europe; 2002.
50. McCarthy OL, Wazwaz O, Calderon VO, Jado I, Saibov S, Stavridis A, et al. Development of an intervention delivered by mobile phone aimed at decreasing unintended pregnancy among young people in three lower middle income countries. BMC Public Health 2018 May 2;18(1):576 [FREE Full text] [doi: 10.1186/s12889-018-5477-7] [Medline: 29716571]
51. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. Annu Rev Public Health 2010;31:399-418. [doi: 10.1146/annurev.publhealth.012809.103604] [Medline: 20070207]
52. Lopez LM, Grey TW, Chen M, Tolley EE, Stockton LL. Theory-based interventions for contraception. Cochrane Database Syst Rev 2016 Nov 23;11:CD007249 [FREE Full text] [doi: 10.1002/14651858.CD007249.pub5] [Medline: 27879980]
53. Lippman SA, Donini A, Díaz J, Chinaglia M, Reingold A, Kerrigan D. Social-environmental factors and protective sexual behavior among sex workers: the Encontros intervention in Brazil. Am J Public Health 2010 Apr 1;100(Suppl 1):S216-S223 [FREE Full text] [doi: 10.2105/AJPH.2008.147462] [Medline: 19762673]
54. L Engle KL, Mwarogo P, Kingola N, Sinkele W, Weiner DH. A randomized controlled trial of a brief intervention to reduce alcohol use among female sex workers in Mombasa, Kenya. J Acquir Immune Defic Syndr 2014 Dec 1;67(4):446-453. [doi: 10.1097/QAI.0000000000000335] [Medline: 25197826]
55. Chernick LS, Stockwell MS, Wu M, Castaño PM, Schnall R, Westhoff CL, et al. Texting to increase contraceptive initiation among adolescents in the emergency department. J Adolesc Health 2017 Dec;61(6):786-790 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.07.021] [Medline: 29056437]
56. McCarthy OL, Zghayyer H, Stavridis A, Adada S, Ahamed I, Leurent B, et al. A randomized controlled trial of an intervention delivered by mobile phone text message to increase the acceptability of effective contraception among young women in Palestine. Trials 2019 Apr 23;20(1):228 [FREE Full text] [doi: 10.1186/s13063-019-3297-4] [Medline: 31014358]
57. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
58. Odeny TA, Newman M, Bukusi EA, McClelland RS, Cohen CR, Camlin CS. Developing content for a mHealth intervention to promote postpartum retention in prevention of mother-to-child HIV transmission programs and early infant diagnosis of HIV: a qualitative study. PLoS One 2014;9(9):e106383 [FREE Full text] [doi: 10.1371/journal.pone.0106383] [Medline: 25181408]
59. Nachega JB, Skinner D, Jennings L, Magidson JF, Altice FL, Burke JG, et al. Acceptability and feasibility of mHealth and community-based directly observed antiretroviral therapy to prevent mother-to-child HIV transmission in South African pregnant women under Option B+: an exploratory study. Patient Prefer Adherence 2016;10:683-690 [FREE Full text] [doi: 10.2147/PPA.S100002] [Medline: 27175068]
60. Ronen K, Unger JA, Drake AL, Perrier T, Akinyi P, Osborn L, et al. SMS messaging to improve ART adherence: perspectives of pregnant HIV-infected women in Kenya on HIV-related message content. AIDS Care 2018 Apr;30(4):500-505 [FREE Full text] [doi: 10.1080/09540121.2017.1417971] [Medline: 29254362]
61. Reiss K, Andersen K, Pearson E, Biswas K, Taleb F, Ngo T, et al. Unintended consequences of mHealth interactive voice messages promoting contraceptive use after menstrual regulation in Bangladesh: intimate partner violence results from a randomized controlled trial. Glob Health Sci Pract 2019 Sep;7(3):386-403 [FREE Full text] [doi: 10.9745/GHSP-D-19-00015] [Medline: 31558596]
62. Jennings L, Gagliardi L. Influence of mHealth interventions on gender relations in developing countries: a systematic literature review. Int J Equity Health 2013 Oct 16;12:85 [FREE Full text] [doi: 10.1186/1475-9276-12-85] [Medline: 24131553]

## Abbreviations

**FSW:** female sex worker
**IUD:** intrauterine device
**LARC:** long-acting reversible contraceptive
**NHMRC:** National Health and Medical Research Council Australia
**RCT:** randomized controlled trial
**SRHR:** sexual and reproductive health and rights
**STI:** sexually transmitted infection
**WHISPER:** Women's Health Intervention using SMS for Preventing Pregnancy

*Edited by G Eysenbach; submitted 19.06.19; peer-reviewed by C Lyons, A Ermias, D Opoku, S Badawy; comments to author 30.09.19; revised version received 24.11.19; accepted 03.02.20; published 29.05.20.*

*Please cite as:*
*Ampt FH, L'Engle K, Lim MSC, Plourde KF, Mangone E, Mukanya CM, Gichangi P, Manguro G, Hellard M, Stoové M, Chersich MF, Jaoko W, Agius PA, Temmerman M, Wangari W, Luchters S*
*A Mobile Phone–Based Sexual and Reproductive Health Intervention for Female Sex Workers in Kenya: Development and Qualitative Study*
*JMIR Mhealth Uhealth 2020;8(5):e15096*
*URL: https://mhealth.jmir.org/2020/5/e15096*
*doi:10.2196/15096*
*PMID:32469326*

Original Paper

# Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the "Needle in a Haystack"

Florian Dittrich[1], Dr Med; Sascha Beck[1,2], Dr Med; Anna Katharina Harren[3], Dr Med; Felix Reinecke[4]; Sebastian Serong[1], Dr Med; Jochen Jung[5], Dr Med; David Alexander Back[6], PD; Milan Wolf[1]; Stefan Landgraeber[1], Prof Dr

[1]Department for Orthopaedics and Orthopaedics Surgery, Saarland University Medical Center and Saarland University Faculty of Medicine, Homburg, Germany

[2]Sportsclinic Hellersen, Lüdenscheid, Germany

[3]Department of Plastic, Reconstructive & Aesthetic Surgery, Specialized Clinic Hornheide, Münster, Germany

[4]Center for Orthopedics and Traumatology, University Medicine Essen, Essen, Germany

[5]Department for Orthopedics, Diakonie Hospital Bad Kreuznach, Bad Kreuznach, Germany

[6]Clinic of Traumatology and Orthopedics, Bundeswehr Hospital Berlin, Berlin, Germany

**Corresponding Author:**
Florian Dittrich, Dr Med
Department for Orthopaedics and Orthopaedics Surgery
Saarland University Medical Center and Saarland University Faculty of Medicine
Kirrbergerstr. 100
Homburg
Germany
Phone: 49 68411624520
Email: florian.dittrich@uks.eu

**Related Article:**

This is a corrected version. See correction statement: https://mhealth.jmir.org/2020/6/e21600/

## Abstract

**Background:** It is undeniable that appropriate smartphone apps offer enormous opportunities for dealing with future challenges in orthopedic surgery and public health, in general. However, it is still unclear how the apps currently available in the two major app stores can be used in daily clinical routine by German orthopedic surgeons.

**Objective:** This study aimed to gain evidence regarding the quantity and quality of apps available in the two major app stores and their suitability for use by orthopedic surgeons in Germany.

**Methods:** We conducted a systematic, keyword-based app store screening to obtain evidence concerning the quantity and quality of commercially available apps. Apps that met the inclusion criteria were evaluated using the *app synopsis–checklist for users* and the German Mobile App Rating Scale for secure use, trustworthiness, and quality.

**Results:** The investigation revealed serious shortcomings regarding legal and medical aspects. Furthermore, most apps turned out to be useless and unsuitable for the clinical field of application (4242/4249, 99.84%). Finally, 7 trustworthy and high-quality apps (7/4249, 0.16%) offering secure usage in the daily clinical routine of orthopedists were identified. These apps mainly focused on education (5/7). None of them were CE (Conformité Européenne) certified. Moreover, there are no studies providing evidence that these apps have any positive use whatsoever.

**Conclusions:** The data obtained in our study suggest that the number of trustworthy and high-quality apps on offer is extremely low. Nowadays, finding appropriate apps in the fast-moving, complex, dynamic, and rudimentarily controlled app stores is most challenging. Promising approaches, for example, systematic app store screenings, app-rating developments, reviews or app libraries, and the creation of consistent standards have been established. However, further efforts are necessary to ensure that these innovative mobile health apps not only provide the correct information but are also safe to use in daily clinical practice.

***(JMIR Mhealth Uhealth 2020;8(5):e17085)*** doi:10.2196/17085

**KEYWORDS**

smartphone; mHealth; app; orthopedics; app store; screening

## Introduction

### Smartphone and Apps

It is only 12 years since smartphones started their triumphant progress through the world of communication media. Nowadays, daily smartphone usage for communication, collection of information, or data for private or professional purposes has become commonplace [1]. The portability and omnipresent accessibility of smartphones enable their usage anywhere and at any time [2]. After initial groundbreaking steps (including the E-Health Act of 2015) [3], the legislature recently gained considerable momentum in the direction of a stringent national digitization strategy. The "Law for better care through digitization and innovation" (Digitale-Versorgung-Gesetz) passed by the Bundestag on November 7, 2019, paved the way for the prescription of apps, the improved use of web video consultation services, and greater data security in the communication of health data [4].

Colloquially known as "apps," mobile apps are defined as application software for mobile-operating systems. They are tailored to the users' individual requirements, bring the smartphone to life, and thus, unleash the full potential of this new technology. These apps usually provide their application-specific functions via an intuitive user interface ("frontend"), specifically adapted to the mobile form factor, and often make use of web resources as well.

It is essential to differentiate between apps developed for patients and those intended for use by medical staff. Apart from apps that provide purely lifestyle advice, there are apps that can directly influence diagnosis or therapy of diseases and, therefore, should be regarded as medical devices [5]. However, there is no standard definition for apps in a medical context, which would enable users to differentiate between "lifestyle apps," "health apps," "medical apps," or "care apps" [6].

Nowadays, apps can also be technically differentiated into native or web apps, each with advantages and disadvantages. Native apps are installed locally on the smartphone and make use of native application programming interfaces, which often leads to a significantly better performance and adaptation to the native look and feel of the respective platform. Web apps are webpages that have been optimized for running on mobile devices. Hybrid apps are based on web technologies but are packed as native apps and, therefore, have an intermediate role.

### App Stores

Apps can be bought and downloaded via several app stores. The major stores are the Google Play Store (Google LLC) and the App Store (Apple Inc). The simple distribution, low development costs, and the ease of use lead to a constantly changing and unmanageable supply. Owing to their complexity and rudimentarily regulated organization, the app stores' offers are nontransparent and heterogeneous [7]. The range is so dynamic that the quantity and quality of apps can vary even from day to day [7,8]. With the rapid development of a fast-moving app industry, the number of apps available in the stores has exploded in the last decade. For instance, the number of apps offered in the "medical" category of the Google Play Store in October 2019 amounted to 42,989 [9].

Owing to inadequate legal, ethical, and medical regulations, many innovative apps operate in a grey area [10]. Recent data scandals have led to a basic distrust of mobile software that could be misused in the context of "big data" [11]. However, comprehensive information on app specifications, which is essential for safe usage in the medical context, is only provided sporadically in the app stores [7,10]. Inadequate store descriptions that provide no transparency and only insufficient information on the intended purpose and limitations of the apps as well as data privacy make it difficult to identify apps designed for the specific requirements of orthopedic surgeons. In addition, specific search terms and keywords are required to find an appropriate app [7]. However, even if there are a large number of matches, only a limited number of results are commonly displayed on the search interfaces provided by the store, and little is known about the criteria and algorithms based on which apps are selected and which of these are listed more prominently [12].

### Mobile Health in Orthopedic Surgeons Clinical Use

Irrespective of rapid developments in the field of medical apps, the information behavior of young physicians has fundamentally changed in recent years [13]. Numerous studies have focused on the potential benefits and consequences of smartphone usage in the fields of orthopedic surgery as a result of the rapidly growing mobile health (mHealth) implementation [14-18]. Currently, there are some app store–based screening reviews of commercially available apps that have been developed to address daily clinical issues in orthopedics [19,20]. These focus on spine surgery [21] or sports medicine [22].

However, to our knowledge, no studies have evaluated the availability and usability of apps directed at the specific needs of German orthopedic surgeons so far. mHealth apps that have not been developed for the German market are dubious from a legal and medical point of view if the algorithms and guidelines used do not comply with German requirements [23].

To address this gap, we conducted a systematic review for quantity and quality of commercially available German apps intended for use in everyday clinical routine in orthopedic surgery practices in Germany.

## Methods

### Systematic App Store–Screening Method

Appropriate apps were identified in a well-established, standardized, keyword-based, and systematic web search in the world's largest web platforms for apps—the Google Play Store and the App Store [24]. The search took place between March 1 and April 27, 2019. For the search in the app stores, 23 German keywords were screened in all categories (Textbox 1).

The search terms, which had a clear relation to the question, were defined by a group of 5 experts before initiating the study.

Two raters independently screened the Google Play Store and the App Store for each individual search term on the day of search. All apps that met the inclusion criteria were placed in individual “My Wish Lists” or Excel spreadsheets (Microsoft Corp).

For the purpose of standardizing the store search, the screening procedure was divided into the following 3 steps (Figure 1):

1. The first step involved a keyword search of apps in the context of orthopedics. The search term “orthopaedic” was always part of the search and was used alone or in combination with the other search terms (Textbox 1) in the Google Play Store and the App Store. The names, icons, and developers of the apps were checked against a priori defined inclusion and exclusion criteria (Figure 1). Apps were excluded if the icon, app name, or developer (1) clearly suggested a game, (2) no German or English name was chosen, and (3) the full use of the app exceeded a price of 5€ (US $5.41) per download.

If an app could not be clearly evaluated based on the overview page, the defined inclusion or exclusion criteria were applied to the store description mentioned in the detailed store view. If it was not possible to clearly differentiate in the detailed store view whether the inclusion criteria were met, the app temporarily remained in the study.

2. All apps deposited by the two examiners in their individual “My Wish List” (Google Play Store) or Excel table (App Store) were then re-evaluated by a 5-member group of experts for the existence of the abovementioned inclusion criteria and entered into single lists per app store. If there was no consensus, the app was still included in the study. If a full and light version of the same app was available, the light version was excluded.

3. If the exclusion criteria were not met, the store descriptions and screenshots were evaluated according to the existence of (1) a German- or English-language store description and (2) orthopedic-specific target group conformity. If the inclusion criteria were met, the apps remained in the study. Apps were excluded if their content was focused on topics that (1) were of no interest to the target group (eg, patients, students, nursing staff, and physiotherapists), (2) were not relevant for clinical orthopedic work (eg, promotion, electronic book [e-book], journal, or congress app), or (3) required external devices for use (eg, accelerometer-based activity monitoring). Further reasons for exclusion were if (4) no developer was identified, (5) no privacy statement was available, or (6) the store description was written in a language other than English or German.

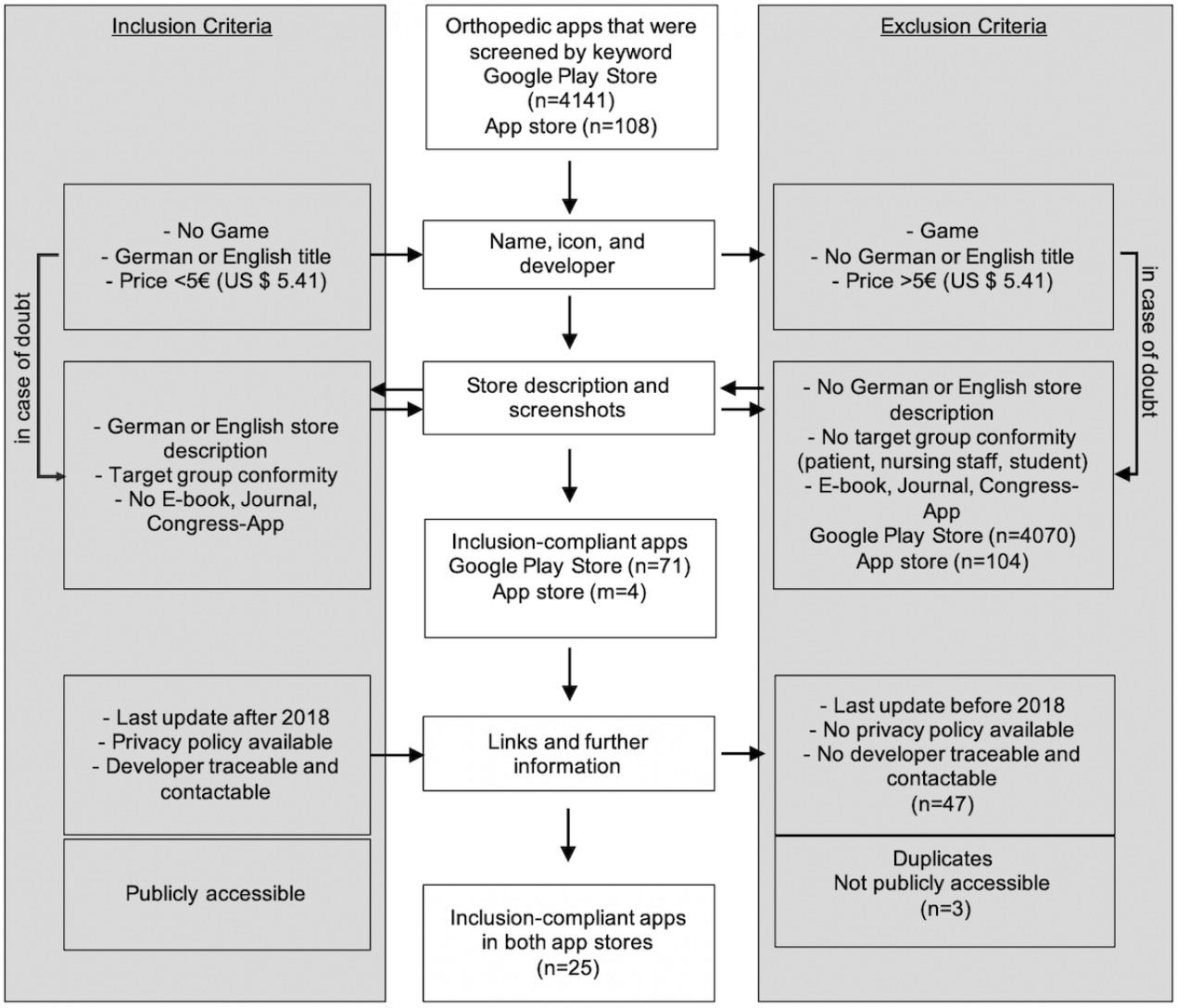
4. The final step tested for (1) the existence of a German- or English-language data protection declaration as well as (2) the identification of the developer and (3) the time of the most recent update (2018 or 2019). Apps that were not updated in at least the previous year were removed.

After duplicates (identical apps found in the Google Play Store and the App Store) had been identified and excluded, the remaining apps, store descriptions, and links to the corresponding app store page were finally merged into one Excel table. If required, we paid for the full version of the app. If there were discrepancies between the two raters, consensus was again reached in the expert group.

**Textbox 1.** Original keywords used in the search and the resulting number of hits in the Google Play Store and App Store.

“Orthopädie” (GPS: 249, AS: 53) OR “Orthopädie” AND “Untersuchung” (GPS: 186, AS: 1) OR “Orthopädie” AND “Untersuchungstechniken” (GPS: 154, AS: 0) OR “Orthopädie” AND “Röntgen” (GPS: 249, AS: 1) OR “Orthopädie” AND “Bildgebung” (GPS: 217, AS: 0) OR “Orthopädie” AND “Operation” (GPS: 249, AS: 0) OR “Orthopädie” AND “Operationstechnik” (GPS: 214, AS: 0) OR “Orthopädie” AND “operativer Zugangsweg” (GPS: 180, AS: 0) OR “Orthopädie” AND “Operationsanleitung” (GPS: 103, AS: 0) OR “Orthopädie” AND “Implantat” (GPS: 248, AS: 0) OR “Orthopädie” AND “Rehabilitation” (GPS: 248, AS: 1) OR “Orthopädie” AND “Nachsorge” (GPS: 124, AS: 0) OR “Orthopädie” AND “Therapie” (GPS: 249, AS: 5) OR “Orthopädie” AND “Diagnose” (GPS: 246, AS: 2) OR “Orthopädie” AND “Diagnostik” (GPS: 247, AS: 1) OR “Orthopädie” AND “Leitlinie” (GPS: 117, AS: 0) OR “Orthopädie” AND “Endoprothetik” (GPS: 122, AS: 9) OR “Orthopädie” AND “Lagerung” (GPS: 111, AS: 0) OR “Orthopädie” AND “Mikrobiologie” (GPS: 122, AS: 0) OR “Orthopädie” AND “Wirbelsäule” (GPS: 161, AS: 20) OR “Orthopädie” AND “Skoliose” (GPS: 106, AS: 2) OR “Orthopädie” AND “Schmerztherapie” (GPS: 112, AS: 13) OR “Orthopädie” AND “Klassifikationen” (GPS: 127, AS: 0)

**Figure 1.** Flowchart screening process. E-book: electronic book.




## Specific App Ranking

Apps that met all inclusion criteria were evaluated using the *app synopsis–checklist for users* [25-27] regarding secure use and trustworthiness in the context of German data protection regulations. Appropriate apps were downloaded, installed, and evaluated from May to July 2019 by 5 raters (FD, DB, KH, FR, and SS) on various smartphones with Android and iPhone operating systems (Samsung Galaxy S8, iPhone 7, and iPhone 8). The evaluators ran all apps on their smartphones for at least 10 days to review all app features and extract data about app features or additional functions.

In a second step, trustworthy apps were rated and ranked using the “German Mobile App Rating Scale” (MARS-G) [28].

All investigations on humans were carried out with the consent of the responsible ethics committee in accordance with the national law and the Declaration of Helsinki of 1975 (in the current, revised version).

## App Synopsis

The app synopsis is a well-established tool for evaluation of the quality and trustworthiness of apps intended for use in Germany [25,26]. It was developed by the Peter L. Reichertz Institute for Medical Informatics at the Hannover Medical School with special focus on the guidelines and regulations applicable in Germany. The *app synopsis* enables app users without a professional information technology background to estimate the trustworthiness of apps. Questions regarding the 8 sections *medical device*, *intended purpose*, *functionality*, *scientific quality*, *restrictions and limits*, *risks*, *reliability of content*, and *data protection* must be marked with one of the following 3 options: “yes,” “no“, or “unclear.” Some, but not all, answer qualities have a higher relevance in the context of an app’s trustworthiness, and the answer options are therefore highlighted based on a signal light system. The better the trustworthiness of an app, the more “green” markers, that is, “yes” answers, it should have obtained. A field highlighted in “red” is an indication for reasonable skepticism about the trustworthiness of an app regarding the respective criterion. If, on the other hand, only fields with a “green” background have been marked, this is an indication of higher trustworthiness compared with apps with fewer or only single positive answers. “Orange” ratings may still indicate trustworthiness for an app, albeit to a lesser extent, and they should be used with caution [27].

### German Mobile App Rating Scale

The MARS-G was developed for professionals to rate app quality and includes the sections *classification*, *quality*, *satisfaction*, and a modifiable app-specific section. The MARS rating is a well-established assessment scale for medical app quality [28]. MARS-G items are scored using a 5-point Likert scale (1=inadequate, 2=poor, 3=acceptable, 4=good, and 5=excellent).

The *classification* section provides descriptive information about the app. The *objective app quality* section includes 19 items divided into 4 subscales, namely *engagement*, *functionality*, *aesthetics*, and *information quality*, and further 10 items comprising 2 subjective subscales, namely *subjective app quality* and *perceived impact*.

The *subjective quality* section contains 4 items evaluating the user's overall satisfaction. The 4 sections of the English MARS version were expanded in the MARS-G by an additional section focusing on the *medical gain* of an app. The 5 subscales and the overall score determine the app's quality [29]. Five reviewers (FD, DB, KH, SS, and FR) watched the associated MARS-G instructional video about how to use the MARS-G scale before rating [30].

### Data Analysis

The paper-based app synopsis and the MARS-G were converted into a digital questionnaire on the *Google Docs* platform (Google LLC). Five reviewers rated the "AO/OTA Fracture Classification" app to evaluate interrater reliability and 8 to 10 randomly selected apps. Data were saved and then transferred to an Excel table. Descriptive statistics were calculated for all items. The interclass correlation coefficients (ICCs) were calculated between the 5 reviewers. We selected an individual absolute agreement intraclass correlation (AA-ICC) for a two-way mixed model on the basis of ICC guidelines by Shrout and Fleiss [31]. The interpretation for ICC interrater agreement measures followed the guidelines of Koo et al [32]. All statistical analyses were conducted using SPSS (version 25, IBM Corp).

## Results

Our systematic web search revealed 4141 hits in the Google Play Store and 108 hits in the App Store using the aforementioned 23 keywords. After evaluating the publicly available information, 1.71% (71/4141) of Google Play Store–screened and 3.7% (4/108) of App Store–screened apps met the formal inclusion criteria. Finally, 0.59% (25/4249) of apps met the minimum requirements of data protection regulation (Figure 1). These apps were downloaded and evaluated using the app synopsis. None of the apps were CE certified. Of these, 8 apps were classified as trustworthy. Good interrater reliability (two-way mixed model single measure AA-ICC=0.78, 95% CI 0.68-0.86) was shown following the guidelines for ICC interpretation established by Koo et al [30]. No trustworthiness markers were missing for the apps *OrthoGuidelines*, *MRI Essentials*, *Touch Surgery: surgical videos*, and *BOSTT*. Another 4 apps, *ICD-10 Diagnoseauskunft*, *DocCheck Help – Arzt*, *AO/OTA Fracture Classification*, and *Calculate by QxMD*, lacked only 1 trust marker on average. Apps that lacked more than one marker on average were classified "not trustworthy." Therefore, concerns had to be raised about the trustworthiness and transparency of the remaining 17 apps that lacked 3.21 (SD 1.26) trustworthiness markers on average (Figure 2).

Subsequently, the quality of the trustworthy apps was evaluated using the MARS-G. The *Calculate by QxMD* app failed to launch at the time of the MARS-G rating and was, therefore, excluded. Moderate interrater reliability (two-way mixed model, single measure AA-ICC=0.58, 95% CI 0.43-0.74) was shown for the MARS-G rating. MARS-G rating revealed the highest overall mean score with 4.3 (SD 0.4) for the app *Touch Surgery: surgical videos* and the lowest score with 3.5 (SD 0.7) for the app *ICD-10 Diagnoseauskunft* (Figures 3 and 4).

**Figure 2.** Mean ratings for trusted apps using the app synopsis—checklist for users. Every single app was ranked by two raters. Apps were primarily ranked in a "trustworthiness scale" (yes=+, unclear=±, and no=−) and secondarily by the following criteria: (1) ascending red (missing), (2) descending green (existing), and (3) ascending orange (unclear) trustworthiness marker. The lower red, higher green, and lower orange markers' quantity, the higher the app's trustworthiness. AO/OTA: Arbeitsgemeinschaft für Osteosynthesefragen/Orthopedic Trauma Association; BoSTT: bone and soft tissue tumors-case studies; ICD-10: International classification of diseases, tenth revision; MRI: magnetic resonance imaging.

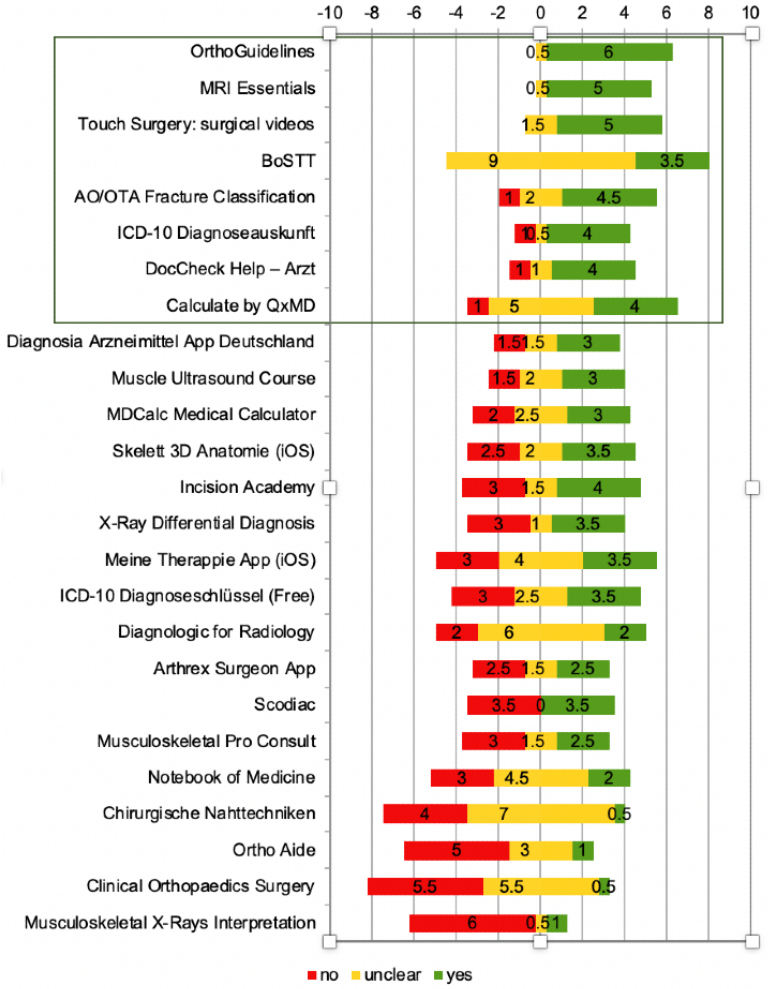



**Figure 3.** Mean overall rating for trusted apps using the "MARS-G". Every single app was ranked by two raters. AO/OTA: Arbeitsgemeinschaft für Osteosynthesefragen/Orthopedic Trauma Association; BoSTT: bone and soft tissue tumors-case studies; ICD-10: International classification of diseases, tenth revision; MRI: magnetic resonance imaging; MARS-G: German Mobile App Rating Scale.

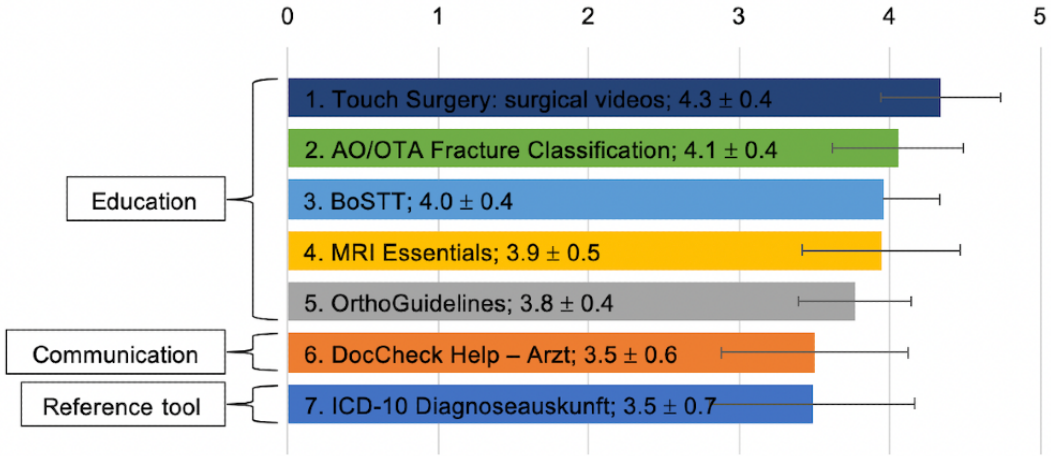
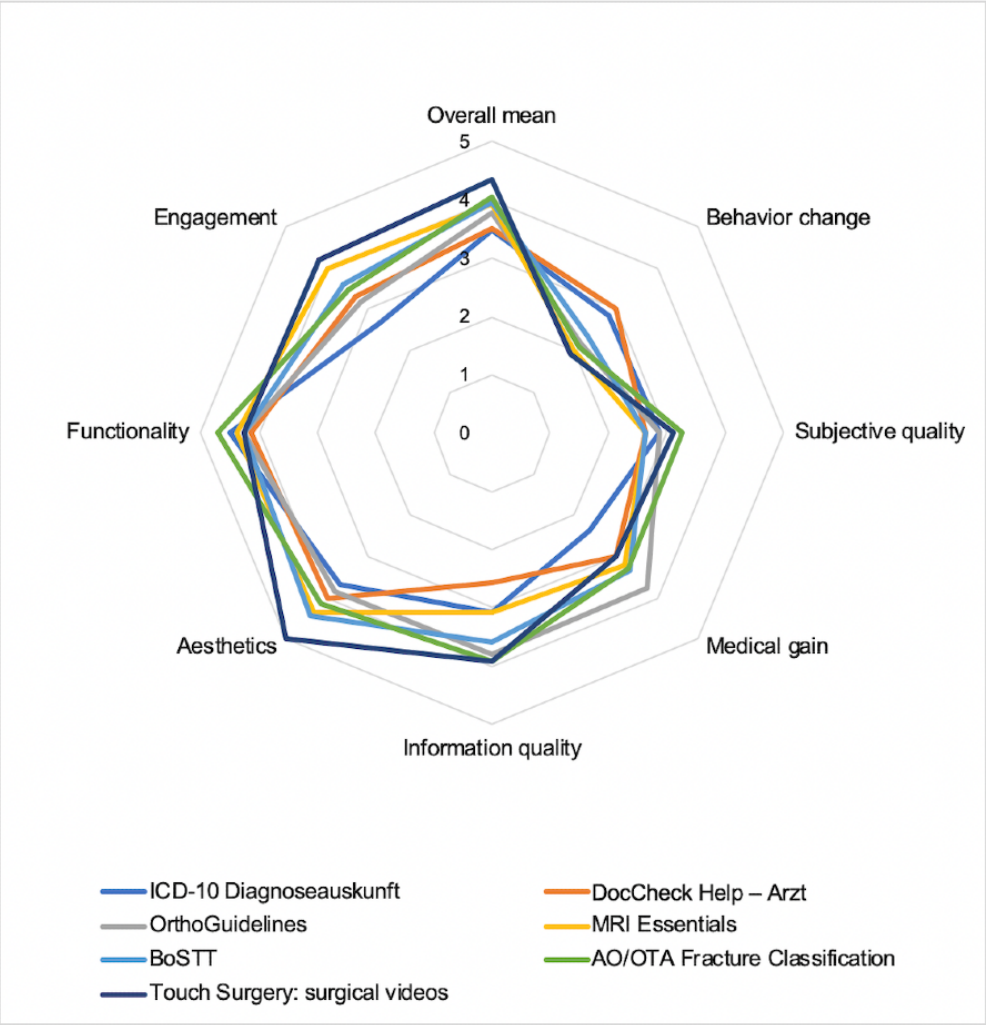


**Figure 4.** Mean section ratings for trusted apps using the "MARS-G". Every single app was ranked by two raters. AO/OTA: Arbeitsgemeinschaft für Osteosynthesefragen/Orthopedic Trauma Association; BoSTT: bone and soft tissue tumors-case studies; ICD-10: International classification of diseases, tenth revision; MRI: magnetic resonance imaging; MARS-G: German Mobile App Rating Scale.




## Discussion

### Principal Findings

The two major app stores were browsed for apps intended for use in everyday clinical routine of orthopedic surgeons. On the basis of keywords, 4249 apps were detected. These were evaluated for quality, safety, and usability. Most of these apps were considered inappropriate for use in daily clinical practice (eg, Games and e-books; 4242/4249, 99.84%). To find an appropriate app, an average of 607 (4249/7) apps had to be screened. Finally, 0.16% (7/4249) of apps were considered reliable, secure, and of high quality in the app synopsis and MARS-G analyses. Interestingly enough, apps that achieved a high score in the app synopsis also received a good evaluation in MARS testing. Nevertheless, none of these apps were CE certified nor had their purpose been evaluated in studies. The remaining apps focused on educational (5/7), communicative (1/7), and reference (1/7) aspects.

The identification of apps tailored to the specific needs of orthopedic surgeons was hampered by a lack of transparency, inadequate store descriptions, missing information on limitations of the app, or lacking precise and public declaration regarding the intended purpose and data protection. Nontransparent store descriptions and missing app-related meta-information represent a great challenge for users when trying to find their way around the app market [12]. Moreover, a systematic evaluation of the top-ranked mental health apps' store descriptions identified the use of scientific language as the most frequently employed strategy to suggest effectiveness [33]. In many cases, the detection of appropriate apps is only successful if specific search terms are used [7]. If the search results in a large number of matches, store providers select a limited number of apps without naming the selection criteria used [12]. This aspect harbors the risk that users will find unsuitable apps and smartphone apps or fail to identify suitable apps, as these may not be displayed at all. One example of inappropriate app usage is the use of *WhatsApp* (Facebook Inc) as a communication tool in everyday clinical practice. Using *WhatsApp* to send patient-related data is not safe and does not comply with the EU-DSGVO regulations because the mobile phone address book is extracted regularly. Messages are already sent in encoded form end-to-end, but the metadata readout is not affected [34].

## Limitations

The presented work has some limitations. The matches using the aforementioned keywords for search in the Google Play Store and the App Store only refer to a priori defined German search words in their nominal form. Moreover, it was not possible to determine how the search result might be influenced by an adjustment of the search terms and the combination of individual words because the underlying app stores' algorithms remain unclear. To ensure a transparent and objective systematic app store search, already well-established methods were applied [24,35,36]. However, none of these methods have been sufficiently validated so far. This must be addressed by future studies.

A further limiting factor is the fact that the apps included in the study were technically and conceptually extremely inhomogeneous, resulting in a great diversity of application areas and legal aspects. Therefore, a discussion is needed as to whether proper evaluation of such a collection of apps is possible with only one standard rating tool. This is underlined by the moderate interrater reliability using MARS-G rating, suggesting that the score itself is rather subjective. These findings are in line with actual systematic app ratings including more than two raters [24,35]. The combination of existing scores might be valuable [36].

## Outlook

Owing to the fast-moving, complex, dynamic, and rudimentarily controlled nature of the app stores, the market is heterogeneous and not transparent for the user on the one hand [7,8], but on the other hand, it might become a highly productive innovation incubator. Therefore, the current situation should be seen both as an opportunity and a risk. Identifying high-quality apps among the wide range of apps currently on the market represents a great challenge. It is like trying to find the famous needle in a haystack. However, some approaches have already addressed this issue:

1. A growing number of publications are critically addressing the currently available apps and have conducted manual systematic app store searches [24,35,36]. These publications may serve as a basis for content and methodological approaches to personal app searches. However, this method is very time consuming. Newly developed semiautomated search methods are based on filtering processes using predefined criteria, for example, the semiautomated retrospective App Store analysis, and might be extended with algorithmic analysis or artificial intelligence in the future [12].
2. By developing the MARS, a first attempt was made to create a tool dedicated to an objective assessment of the app's content and technical specifications, which is essential to enable a comparison between apps [28]. The app synopsis, also used in this study, primarily focuses on an app's trustworthiness. Nevertheless, apps that collect and process sensitive patient-related data must fulfill higher data protection requirements than apps that are used for coding purposes and do not collect data at all. But evaluation with a standard-based tool might lead to a false-negative rating of the coding app. Therefore, the existing ratings have been constantly improved, and new, more specific tools have been developed [29,37]. An increase in app rating quality might be achieved if adaptive ratings focusing on the intended app purpose were developed.
3. Several professional associations as well as private institutions aim to review apps and publish them in app libraries. The
   NHS Apps Library
   only recommends safe and secure apps in the United Kingdom. Developers must answer a standardized, transparent, and publicly available range of digital assessment questions designed by experts from technical and policy backgrounds [38]. For mental health apps, a nonprofit organization, in cooperation with several universities, provides guidelines and app reviews on the webpage PsyberGuide. The standardized review process is based on credibility, user experience, and transparency using established rating tools [39]. In the field of orthopedic surgery, the private webpage
   TopOrthoApps
   gives app information and reviews, though the review process is not transparent and the studies presented seem outdated [40].

In the absence of consistent legal, ethical, and medical regulations, numerous innovative apps remain in a grey area and struggle to deploy their full potential. In times when international technology concerns are already optimizing innovative technologies (eg, the use of artificial intelligence) in apps, the framework conditions for a solid but also dynamic and adaptive mHealth strategy must be developed in the German health care system.

## Conclusions

The benefits of the appropriate use of smartphones and apps in the field of orthopedic surgery are undeniable and have enormous potential for dealing with future challenges in public health. The data gained in our study suggest that the number of trustworthy and high-quality apps on offer is extremely low. Most of the apps display serious shortcomings regarding legal and medical aspects. The fast-moving, complex, and dynamic nature of the app stores, which are under only rudimentary control, harbors the risk of inappropriate app usage. However, the stores also provide important innovations in health care. The search for the appropriate app is a considerable challenge. Promising approaches, for example, systematic app store screenings, app rating developments, reviews or app libraries, and the creation of consistent standards have already been established. Further efforts and interdisciplinary cooperation are required to detect innovative mHealth solutions that can be utilized in a safe and secure manner in the work of orthopedic surgeons.

## Acknowledgments

The authors acknowledge support by the Deutsche Forschungsgemeinschaft (German Research Foundation) and Saarland University within the funding program Open Access Publishing.

## Conflicts of Interest

None declared.

## References

1. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
2. Cha SS, Seo BK. Smartphone use and smartphone addiction in middle school students in Korea: Prevalence, social networking service, and game use. Health Psychol Open 2018;5(1):2055102918755046 [FREE Full text] [doi: 10.1177/2055102918755046] [Medline: 29435355]
3. Bundesministerium für Gesundheit. 2015. E-Health-Gesetz verabschiedet [E-health law passed] URL: https://www.bundesgesundheitsministerium.de/ministerium/meldungen/2015/e-health.html [accessed 2019-12-29]
4. Bundesministerium für Gesundheit. 2020. Ärzte sollen Apps verschreiben können Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz - DVG) URL: https://www.bundesgesundheitsministerium.de/digitale-versorgung-gesetz.html [accessed 2019-12-28]
5. Nasser FB, Trevena L. There's an app for that: a guide for healthcare practitioners and researchers on smartphone technology. Online J Public Health Inform 2015;7(2):e218 [FREE Full text] [doi: 10.5210/ojphi.v7i2.5522] [Medline: 26392848]
6. Albrecht UV. Transparency of health-apps for trust and decision making. J Med Internet Res 2013 Dec 30;15(12):e277 [FREE Full text] [doi: 10.2196/jmir.2981] [Medline: 24449711]
7. Dittrich F, Beck S, Busch A, Dudda M, Harren A, Sander A, et al. [Is the discharge letter still relevant? : Chances and risks of 'Medical apps' in orthopedics and traumatology]. Orthopade 2018 Oct;47(10):842-848. [doi: 10.1007/s00132-018-3598-4] [Medline: 30039468]
8. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. JMIR Mhealth Uhealth 2016 Aug 9;4(3):e96 [FREE Full text] [doi: 10.2196/mhealth.6020] [Medline: 27507641]
9. AppBrain. 2019. Most Popular Google Play Categories URL: https://www.appbrain.com/stats/android-market-app-categories [accessed 2019-05-10] [WebCite Cache ID 78GbEu9n9]
10. Albrecht UV, Hillebrand U, von Jan U. Relevance of trust marks and CE labels in German-language store descriptions of health apps: analysis. JMIR Mhealth Uhealth 2018 Apr 25;6(4):e10394 [FREE Full text] [doi: 10.2196/10394] [Medline: 29695374]
11. Stroud C. Forbes. 2018 Apr 30. Cambridge Analytica: The Turning Point In The Crisis About Big Data URL: https://www.forbes.com/sites/courtstroud/2018/04/30/cambridge-analytica-the-turning-point-in-the-crisis-about-big-data/ [accessed 2019-05-10] [WebCite Cache ID 78Gc6v0IQ]
12. Albrecht UV, Hasenfuß G, von Jan U. Description of cardiological apps from the German app store: semiautomated retrospective app store analysis. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11753 [FREE Full text] [doi: 10.2196/11753] [Medline: 30459140]
13. Scott K, Morris A, Marais B. Medical student use of digital learning resources. Clin Teach 2018 Feb;15(1):29-33. [doi: 10.1111/tct.12630] [Medline: 28300343]
14. Duncan SF, Hendawi TK, Sperling J, Kakinoki R, Hartsock L. iPhone and iPad use in orthopedic surgery. Ochsner J 2015;15(1):52-57 [FREE Full text] [Medline: 25829881]
15. Sahoo S, Mohr J, Strnad GJ, Vega J, Jones M, Schickendantz MS, et al. Validity and efficiency of a smartphone-based electronic data collection tool for operative data in rotator cuff repair. J Shoulder Elbow Surg 2019 Jul;28(7):1249-1256. [doi: 10.1016/j.jse.2018.12.009] [Medline: 31056396]
16. Pongkunakorn A, Chatmaitri S, Diewwattanawiwat K. Use of smartphone to improve acetabular component positioning in total hip athroplasty: A comparative clinical study. J Orthop Surg (Hong Kong) 2019;27(1):2309499019825578. [doi: 10.1177/2309499019825578] [Medline: 30798733]
17. Goldstein Y, Schermann H, Dolkart O, Kazum E, Rabin A, Maman E, et al. Video examination via the smartphone: A reliable tool for shoulder function assessment using the constant score. J Orthop Sci 2019 Sep;24(5):812-816. [doi: 10.1016/j.jos.2018.12.023] [Medline: 30686690]
18. Longoni L, Brunati R, Sale P, Casale R, Ronconi G, Ferriero G. Smartphone applications validated for joint angle measurement: a systematic review. Int J Rehabil Res 2019 Mar;42(1):11-19. [doi: 10.1097/MRR.0000000000000332] [Medline: 30640272]
19. Franko OI. Smartphone apps for orthopaedic surgeons. Clin Orthop Relat Res 2011 Jul;469(7):2042-2048 [FREE Full text] [doi: 10.1007/s11999-011-1904-0] [Medline: 21547414]

20. Datillo JR, Gittings DJ, Sloan M, Hardaker WM, Deasey MJ, Sheth NP. 'Is There An App For That?' Orthopaedic patient preferences for a smartphone application. Appl Clin Inform 2017 Aug 16;8(3):832-844 [FREE Full text] [doi: 10.4338/ACI-2017-04-RA-0058] [Medline: 28885637]
21. Robertson GA, Wong SJ, Brady RR, Subramanian AS. Smartphone apps for spinal surgery: is technology good or evil? Eur Spine J 2016 May;25(5):1355-1362. [doi: 10.1007/s00586-015-3932-z] [Medline: 25870076]
22. Wong SJ, Robertson GA, Connor KL, Brady RR, Wood AM. Smartphone apps for orthopaedic sports medicine - a smart move? BMC Sports Sci Med Rehabil 2015;7:23 [FREE Full text] [doi: 10.1186/s13102-015-0017-6] [Medline: 26464800]
23. Albrecht UV, Pramann O. Haftungsfragen beim Einsatz von Gesundheits-Apps: Nobody is perfect. Dtsch Arztebl 2018;115(12):520 [FREE Full text]
24. Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2016 Jun 14;4(2):e74 [FREE Full text] [doi: 10.2196/mhealth.5882] [Medline: 27302310]
25. Albrecht UV, Noll C, von Jan U. App-synopsis: self-assessment on trust or distrust of health-apps. Stud Health Technol Inform 2014;202:233-236. [Medline: 25000059]
26. Albrecht UV, von Jan U, Jungnickel T, Pramann O. App-synopsis - standard reporting for medical apps. Stud Health Technol Inform 2013;192:1154. [Medline: 23920928]
27. Albrecht UV. PLRI MedAppLab.: PLRI MedAppLab App-Synopsis - User- Deutsche Version URL: https://plrimedapplab.weebly.com/uploads/7/4/0/7/7407163/app-synopse-hersteller-dt-040315-1911.pdf [accessed 2020-02-19]
28. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
29. Messner ET, Barke A, Baumeister H, Stoyanov S, Hides L, Kavanagh D, et al. Development and validation of the German version of the mobile application rating scale (MARS-G). JMIR mHealth and uHealth;JMIR Preprints 2019 preprint. [doi: 10.2196/14479]
30. MARS-Rating Universität Ulm. YouTube. 2017. MARS - Mobile Anwendungen Rating Skala URL: https://www.youtube.com/watch?v=5vwMiCWC0Sc&t=400s [accessed 2020-02-19]
31. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [doi: 10.1037//0033-2909.86.2.420] [Medline: 18839484]
32. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163 [FREE Full text] [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
33. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019;2:18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
34. ÄrzteZeitung. 2018 May 29. WhatsApp für Ärzte dienstlich ab sofort tabu URL: https://www.aerztezeitung.de/praxis_wirtschaft/special-arzt-patient/article/964884/datenschutz-whatsapp-aerzte-dienstlich-ab-sofort-tabu.html [accessed 2020-02-19]
35. Knitza J, Tascilar K, Messner E, Meyer M, Vossen D, Pulla A, et al. German mobile apps in rheumatology: review and analysis using the mobile application rating scale (MARS). JMIR Mhealth Uhealth 2019 Aug 5;7(8):e14991 [FREE Full text] [doi: 10.2196/14991] [Medline: 31381501]
36. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 4;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
37. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
38. National Health Service. 2019. NHS Apps Library URL: https://www.nhs.uk/apps-library/ [accessed 2020-02-19]
39. PsyberGuide | A Project of One Mind. 2019. URL: https://psyberguide.org [accessed 2020-02-11]
40. TopOrthoApps. URL: http://toporthoapps.com [accessed 2020-02-11]


## Abbreviations

**AA-ICC:** absolute agreement intraclass correlation
**ICC:** interclass correlation coefficients
**MARS-G:** German Mobile App Rating Scale
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 16.11.19; peer-reviewed by J Knitza, E Sadeghi-Demneh; comments to author 09.01.20; revised version received 28.01.20; accepted 30.01.20; published 07.05.20.*

*Please cite as:*
*Dittrich F, Beck S, Harren AK, Reinecke F, Serong S, Jung J, Back DA, Wolf M, Landgraeber S*
*Analysis of Secure Apps for Daily Clinical Use by German Orthopedic Surgeons: Searching for the "Needle in a Haystack"*
*JMIR Mhealth Uhealth 2020;8(5):e17085*
*URL: https://mhealth.jmir.org/2020/5/e17085*
*doi:10.2196/17085*
*PMID:32379054*

Original Paper

# Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review

Nancy Lau[1,2], PhD; Alison O'Daffer[2], BA; Susannah Colt[2], MA; Joyce P Yi-Frazier[2], PhD; Tonya M Palermo[3,4], PhD; Elizabeth McCauley[1,4], PhD; Abby R Rosenberg[2,5,6], MD, MS, MA

[1]Department of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, WA, United States

[2]Palliative Care and Resilience Lab, Center for Clinical and Translational Research, Seattle Children's Research Institute, Seattle, WA, United States

[3]Department of Anesthesiology and Pain Medicine, University of Washington School of Medicine, Seattle, WA, United States

[4]Center for Child Health, Behavior, and Development, Seattle Children's Research Institute, Seattle, WA, United States

[5]Cambia Palliative Care Center of Excellence, University of Washington, Seattle, WA, United States

[6]Department of Pediatrics, University of Washington School of Medicine, Seattle, WA, United States

**Corresponding Author:**
Nancy Lau, PhD
Department of Psychiatry and Behavioral Sciences
University of Washington School of Medicine
1900 Ninth Ave
JMB 10-C
Seattle, WA, 98101
United States
Phone: 1 2068840569
Email: nancy.lau@seattlechildrens.org

## Abstract

**Background:** In an oversaturated market of publicly available mobile apps for psychosocial self-care and stress management, health care providers, patients, and consumers interested in mental health–related apps may wonder which, if any, are efficacious. Readily available metrics for consumers include user popularity and media buzz rather than scientific evidence.

**Objective:** This systematic review aimed to (1) examine the breadth of therapeutic contents and features of psychosocial wellness and stress management apps available to self-help seekers for public download and (2) determine which of these apps have original research support.

**Methods:** First, we conducted a systematic review of publicly available apps on the iPhone App Store (Apple Inc) and Android Google Play (Google LLC) platforms using conventional self-help-seeking search terms related to wellness and stress. The results were limited to English-language apps available for free download. In total, 2 reviewers independently evaluated all apps and discussed the findings to reach 100% consensus regarding inclusion. Second, a literature review was conducted on the included apps to identify supporting studies with original data collection.

**Results:** We screened 3287 apps and found 1009 psychosocial wellness and stress management apps. Content varied widely. The most common evidence-based strategy was mindfulness-meditation, followed by positive psychology and goal setting. Most apps were intended to be used as self-help interventions, with only 1.09% (11/1009) involving an electronic therapist and 1.88% (19/1009) designed as a supplement to in-person psychotherapy. Only 4.66% (47/1009) of apps targeted individuals with psychological disorders, and less than 1% of apps (6/1009, 0.59%) targeted individuals with other chronic illnesses. Approximately 2% (21/1009, 2.08%) were supported by original research publications, with a total of 25 efficacy studies and 10 feasibility studies. The *Headspace* mindfulness app had the most evidence, including 8 efficacy studies. Most other scientifically backed apps were supported by a single feasibility or efficacy study.

**Conclusions:** Only 2.08% (21/1009) of publicly available psychosocial wellness and stress management mobile apps discoverable to self-help seekers have published, peer-reviewed evidence of feasibility and/or efficacy. Clinicians and investigators may use these findings to help patients and families navigate the volume of emerging digital health interventions for stress management and wellness.

*(JMIR Mhealth Uhealth 2020;8(5):e17798)* doi:10.2196/17798

**KEYWORDS**

mHealth; mobile health; mental health

## Introduction

### Background

Within the past decade, smartphones have become ubiquitous in personal, social, and work life [1], irrespective of gender, race and ethnicity, and socioeconomic status [2]. Overall, 75% of Americans own a smartphone, and 83% of them never leave home without it [1,3]. On average, a person checks his or her phone 150 times per day [4]. Owing to the pervasiveness of smartphones in modern day culture, technological innovations may be leveraged to disseminate *in the moment* behavioral change interventions designed to promote healthy behaviors [5]. There is a robust market for health apps, with 325,000 available for download as of 2017 and a growth rate of 25% year to year [6]. More than half of mobile phone users have downloaded a health-related mobile app, and the pace of development of evidence-based apps tested in research settings has lagged far behind than that of the commercial sector [7-9].

In particular, mobile health (mHealth) apps focused on promoting emotional health and adaptive coping have become increasingly popular. Mental health symptoms such as anxiety and stress are prevalent and disruptive. Overall, 75% of adults in the United States report significant stress, and 19% have mental health disorders [10]. Anxiety disorders impact up to 30% of individuals worldwide, leading to severe societal and economic burden [11]. Work-related stress alone costs the US economy US $402 billion [12]. Disseminating psychosocial interventions via mHealth technologies confers the advantage of universal accessibility regardless of geographic and economic restrictions [1,13]. According to the US National Comorbidity Survey (a nationally representative large-scale mental health study), common barriers to seeking mental health care include financial constraints, stigma, and a desire for self-management of symptoms [14]. In other studies, most individuals reported interest in using a mobile app for self-management of anxiety, stress, and depression if services were available for free [8,15]. In total, two recent meta-analyses of randomized controlled trials (RCTs) showed that mHealth interventions for anxiety and depression showed small positive effects when compared with an active control condition [11,16].

Despite the high demand and potential advantages of these apps, there is a lack of quality control standards or readily accessible information to consumers on whether or which apps work in an oversaturated market. Thus, leveraging mHealth technologies brings both benefits and new challenges. In efforts to review publicly available apps using a direct-to-consumer approach, recent mHealth reviews have used a search strategy that involves entering key terms directly into the search engines of mobile app platforms [17-20]. In a review of iOS App Store mobile apps on Apple devices (Apple Inc), researchers identified 60 mobile apps that delivered at least one evidence-based stress management strategy (eg, mindfulness, progressive muscle relaxation, and biofeedback) [19]. A recent review of apps for depression, anxiety, posttraumatic stress disorder (PTSD), and alcohol use found that *evidence-based mobile apps* (ie, apps tested via formal research methods and published in the scientific literature) are often unavailable for download to the general public; in addition, apps available to consumers on commercial platforms are highly variable with regard to the inclusion of *evidence-based content* (ie, content derived from empirically supported therapeutic approaches) [20].

### Objectives

With an overabundance of publicly available apps for stress management and psychosocial self-care, consumers may struggle with a paradox of choice, regardless of whether they are providers seeking to make app recommendations, patients seeking additional mental health support, or app-savvy digital natives interested in self-help. Readily available metrics are app visibility because of ranked lists, user popularity, media buzz, and user satisfaction ratings. When an app purports to be based in science, its *scientific backing* may not reach the classical standards of research rigor. It remains unclear whether popular apps that consumers gravitate toward *work*. In this study, we broadly reviewed popular apps available to all manner of consumers (ie, the general public, patients, and providers) for free download and their treatment content, user ratings, costs, and the evidence base in support of them. We presented findings from a review of 1009 publicly available mobile apps on Apple Store and Google Play (Google LLC) platforms using common self-help-seeking search terms for psychosocial wellness and stress management. Our review spanned Apple and Android devices that together represent 99% of the smartphone user market; 54.4% of US smartphone owners use Android devices and 44.3% use Apple devices [21]. After systematically searching both mobile app platforms (step 1), we supplemented this direct-to-consumer approach by conducting a literature review of the apps identified (step 2).

Our research questions were as follows: (1) What are the active therapeutic components and features of publicly available psychosocial wellness, coping, and stress management mobile apps? (2) Do any of these mobile apps have evidence in support of their feasibility/acceptability or efficacy in the published scientific literature? We hypothesized that the majority of consumer apps identified would not contain evidence-based therapeutic strategies, and even fewer would have published research supporting the apps themselves. We translated research findings to clinical practice by describing the breadth of popular wellness apps for stress management and by identifying the few apps with scientific backing.

## Methods

### Searching the Apple Store and Google Play: Step 1

We systematically identified and evaluated apps using a modified version of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines [22]; adjustments were

made because of the differing methodology of directly searching app store platforms. Our search strategy protocol is available in Multimedia Appendix 1.

We searched the mobile app platforms App Store iOS (Apple Inc) and Google Play in September 2018. Inclusion criteria were (1) a focus on stress management and/or psychosocial wellness, (2) available in English, and (3) free for download (including those with free downloads for *basic* subscriptions, with additional fees for extra features).

Specifically, we first created a list of conventional self-help-seeking search terms from mental health and positive psychology background literature [23-31]; we refined the list in discussions among our interdisciplinary research team, which includes intervention science researchers, health services researchers, physicians, social workers, and psychologists. Then, we entered into Apple Store and Google Play search engines the 14 conventional self-help-seeking search terms agreed upon by our team: stress, resilience, goal setting, relaxation, mindfulness, mood, coping, gratitude, optimism, hope, happiness, sadness, self-compassion, and self-care. We noted that the app results were displayed on both search engines in the order of popularity, based on proprietary algorithms. Hence, we screened the first 100 apps populated for each search term. Indeed, research suggests that smartphone users limit their searches to the first page of results (which contains 10 apps) [32], so screening the first 100 apps was deemed sufficient. In addition, we screened the popularity lists in *health and fitness* and *kids and family* categories for apps that met the inclusion criteria (these popularity lists are displayed on Apple and Google Play platforms). Two authors (NL and AO) independently reviewed all apps for inclusion and discussed the findings to reach 100% consensus. Duplicates were removed.

### Data Extraction Procedures

For each app that met the inclusion criteria, 2 authors (NL and AO) extracted the following App Store iOS (Apple Inc) and Google Play product page data from November 2018 to February 2019 and discussed findings to reach 100% consensus: intervention and didactic content, target audience, and whether there were in-app paid features. In creating our database of intervention and didactic content, we used an all-inclusive approach to delineate content categories. For example, if an app description included meditation, mood tracking, artificial intelligence, and chat forums, we created unique content categories for each. Our goal was to provide a comprehensive representation of all intervention and didactic content as described on product pages by the app developers. To do so, we iteratively coded apps in sets of 50 and expanded the number of content categories as needed until there were no new content categories that arose. This resulted in the final version of the database, which contained 31 unique intervention and didactic content categories. During consensus conversations in March 2019 to April 2019, our process was to re-review App Store iOS (Apple Inc) and Google Play product page data to resolve discrepancies.

### Literature Review: Step 2

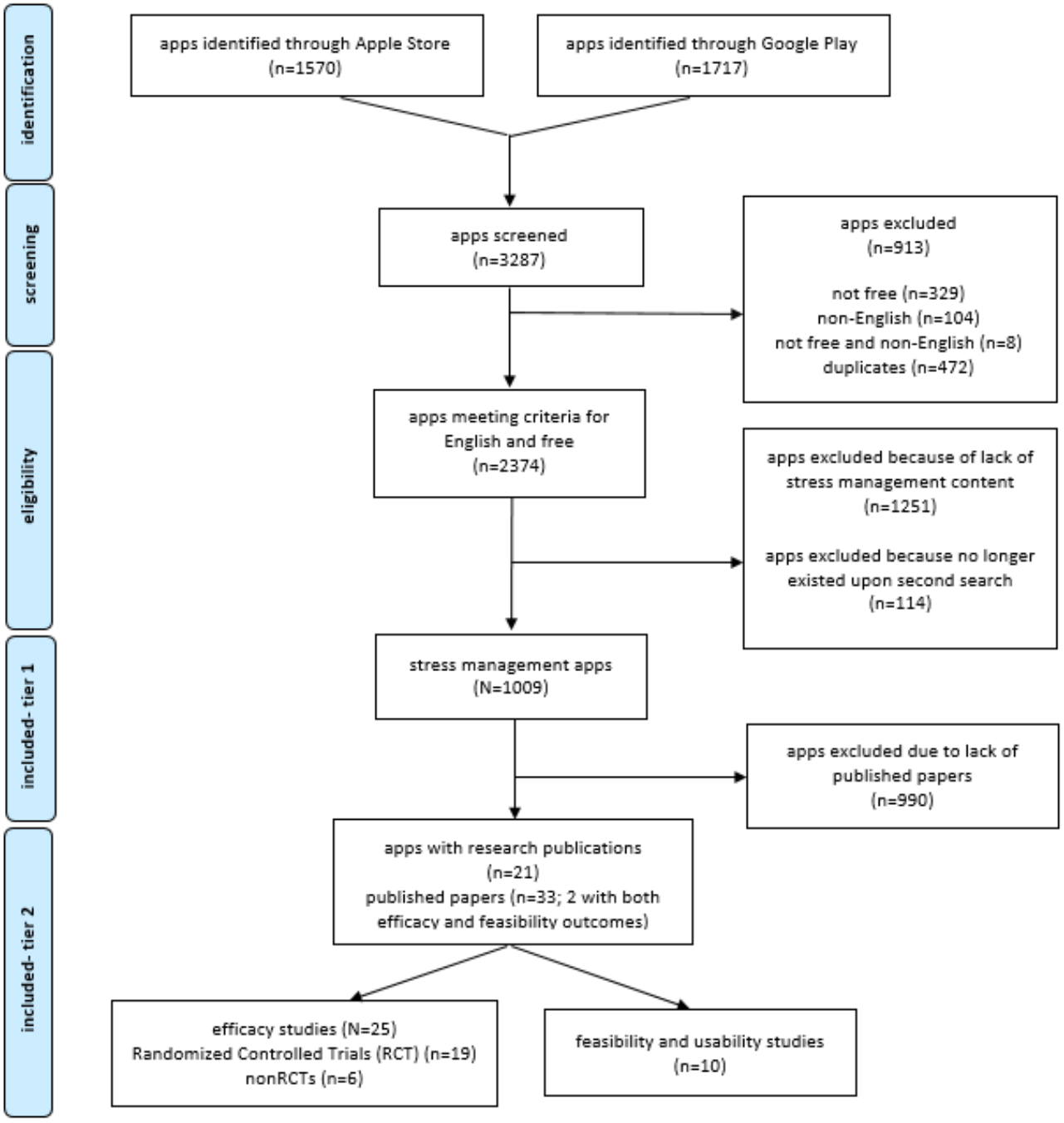
After all mobile apps were identified in step 1, 2 authors (AO and SC) conducted a literature review via Google Scholar, MEDLINE, and PsycINFO databases using the search terms *[app name]* AND *smartphone* from April 2019 to June 2019 to identify peer-reviewed papers supporting each of the identified apps. Furthermore, 2 authors (NL and AO) retrieved and independently reviewed the full text of all eligible studies to extract relevant feasibility and efficacy outcomes. We included research papers published in peer-reviewed journals and in English and included qualitative and/or quantitative studies with original data collection. We excluded conference presentations, editorials, commentaries, and study protocols. In consultation with a medical librarian, we chose to exclude nonpeer-reviewed scholarly works before publication in a peer-reviewed journal because the information included in conference abstracts lacks the rigor and external validity inherent in peer review. From the 33 included papers, we retrieved the following information: participants, sample size, study type, treatment conditions, and outcomes reported. Two authors (NL and AO) independently assessed study quality using the Cochrane collaboration's tool for assessing the risk of bias [33], evaluating random sequence generation (selection bias), allocation concealment (selection bias), blinding of participants and personnel (performance bias), blinding of outcome assessment (detection bias), incomplete outcome data (attrition bias), selective reporting (reporting bias), and other bias. We coded each category as low, high, or unclear risk of bias according to established standards. We resolved minor discrepancies in coding by referring to the journal papers themselves.

## Results

### Searching the Apple Store and Google Play: Step 1

We screened a total of 3287 apps (Figure 1). Of 3287 apps, 913 (27.78%) were excluded after the initial screening process because they were not available for free download, not in English, or were duplicates. Of the remaining 2374 apps, 1251 (52.70%) were excluded because they did not contain psychosocial wellness or stress management content (eg, health and fitness apps for exercise, nutrition, and weight loss). Of the remaining 1123 apps, 114 (10.15% that we initially found on the Apple Store or Google Play no longer existed 3 months later when authors attempted to refer to the original source for consensus conversations in March 2019 to April 2019. We ultimately included 1009 apps in this review.

**Figure 1.** Preferred reporting items for systematic reviews and meta-analyses diagram.




### *Characteristics of All Included Apps (N=1009)*

For the pooled 1009 apps, we found 31 unique intervention and didactic content categories (eg, cognitive behavioral therapy, mindfulness-meditation, and journaling; Figure 2). Emotional-inspirational quotes were the most common app component, included in 22.99% (232/1009) of apps. Other common components included in 15% of apps or more were goal setting, positive psychology, journaling, music, mindfulness-meditation, and educational materials. Only 4.66% (47/1009) of apps were designed specifically for psychological disorders, less than 1% of apps were designed for chronic illnesses (6/1009, 0.59%), and 3.96% (40/1009) of apps were designed for youths and/or young adults. Most apps were intended to be used as self-help interventions, with only 1.09% (11/1009) involving an electronic therapist (e-therapist) and 1.88% (19/1009) designed as a supplement to in-person psychotherapy.

## Literature Review: Step 2

### *Characteristics of Subset of Apps With Research Support (n=21)*

We found supporting original research publications for 2.08% of apps (21/1009 identified). For this subset of apps, the most common therapeutic component was mindfulness-meditation, an evidence-based treatment strategy that was incorporated into 67% (14/21) of apps with published research followed closely by mood and symptom monitoring. All other common app features (included in ≥15% of apps) were also evidence-based

treatment strategies: cognitive behavioral therapy, positive psychology, and relaxation (Multimedia Appendix 2 and Figure 3). For each app, the average user satisfaction ratings and the number of user ratings varied widely (Multimedia Appendix 3).

### *Peer Review Publications*

A total of 33 peer-reviewed papers supported the 21 apps; 23 of these papers were efficacy studies [34-58], 8 were feasibility or usability studies, and 2 were combined efficacy and feasibility studies [39,59]. For each of the 21 apps with research support, the number of associated peer-reviewed publications ranged from 1 to 8 (*Headspace* [46,48-54]; Multimedia Appendix 3). The majority of apps (16/21, 76%) only had 1 publication [34,35,37,38,43,44,46,55,56,58-62]. *10% Happier*, *Calm*, and *Headspace* were the only research-supported apps we found on ranked *health and fitness* popularity lists.

Of the 25 efficacy studies, 19 were RCTs published between 2015 and 2019 [34-37,40-42,44-46,48,49,51-56,58,59], with 6 of the 19 trials testing *Headspace* [46,48-54]. The majority of studies used samples of convenience, that is, college students [34,35,40,46,51,57] or users who had already downloaded the app [37,39,52,55,56,58]. Sample sizes for efficacy studies ranged from 19 [48] to 153,834 [39]. Treatment duration ranged from a single session of self-directed app use [39] to 6 months [38]. All 16 apps with peer-reviewed publications that reported app efficacy showed some evidence of improving psychosocial outcomes over time (Multimedia Appendix 4). Studies collected varying outcome measures, ranging from unstructured self-directed app use [41-45,50,62] to providing a sequential program of set frequency and duration [49,53]. In a subset of studies where effect sizes were reported; treatment effects ranged from small to large (Multimedia Appendix 4) [41,50,53,58]. Of the 10 apps with feasibility or usability studies, the majority (8/10 apps) reported that users found the apps to be enjoyable, accessible, and acceptable (Multimedia Appendix 5) [39,59-67]. Sample sizes for feasibility and usability studies ranged from 1 [60] to 1255 [64].

### *Risk of Bias Assessment*

The risk of bias was evaluated for all 25 efficacy studies (Figure 4). Of the 19 RCTs, 18 reported random sequence generation and allocation concealment. For the blinding of participants and personnel domain and the outcome assessment domain, studies were roughly split in half between high and low risk; high-risk studies consisted of study designs with no control group, a waitlist control group, or an educational handout control group. For selective reporting bias, 6 were considered low risk, 1 high risk, and 18 were unclear. For other biases, 12 were considered low risk, 12 were high risk, and 1 was unclear.

**Figure 2.** Intervention and didactic content for all stress management apps (N=1009). Content categories were assigned based on descriptions by the app developer. Categories were not mutually exclusive, and a single app could be represented across one or more. E-therapist: electronic therapist.

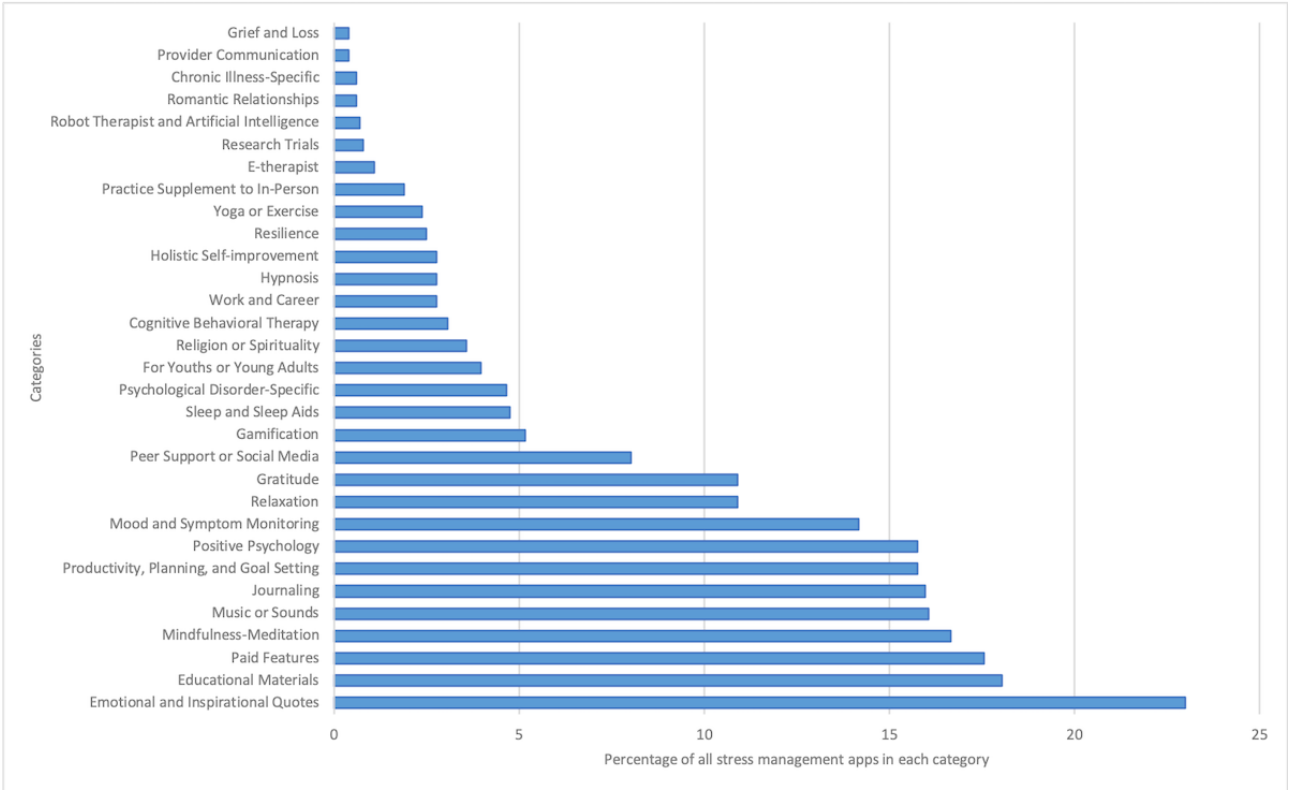


**Figure 3.** Intervention and didactic content for research-supported apps (n=21) vs all stress management apps (N=1009). Content categories were assigned based on descriptions by the app developer. Categories were not mutually exclusive, and a single app could be represented across one or more.

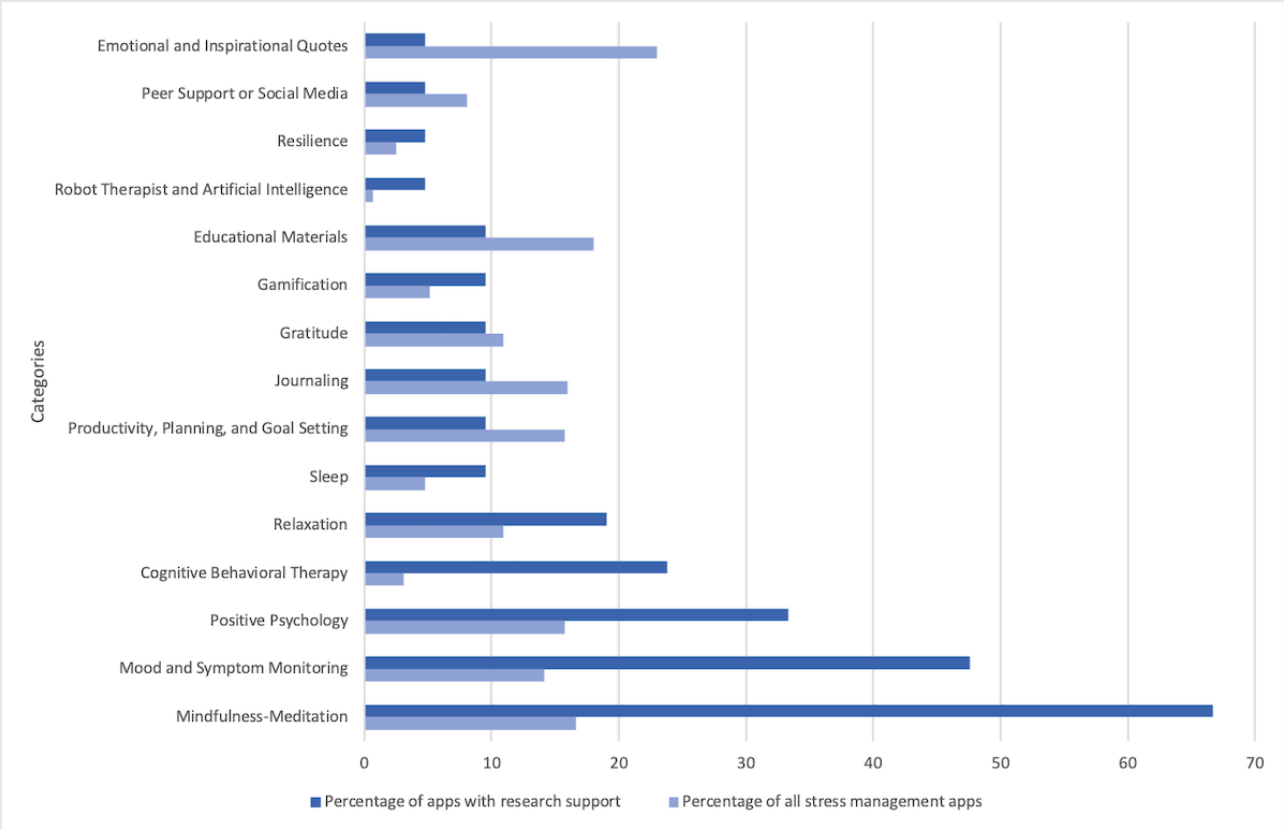
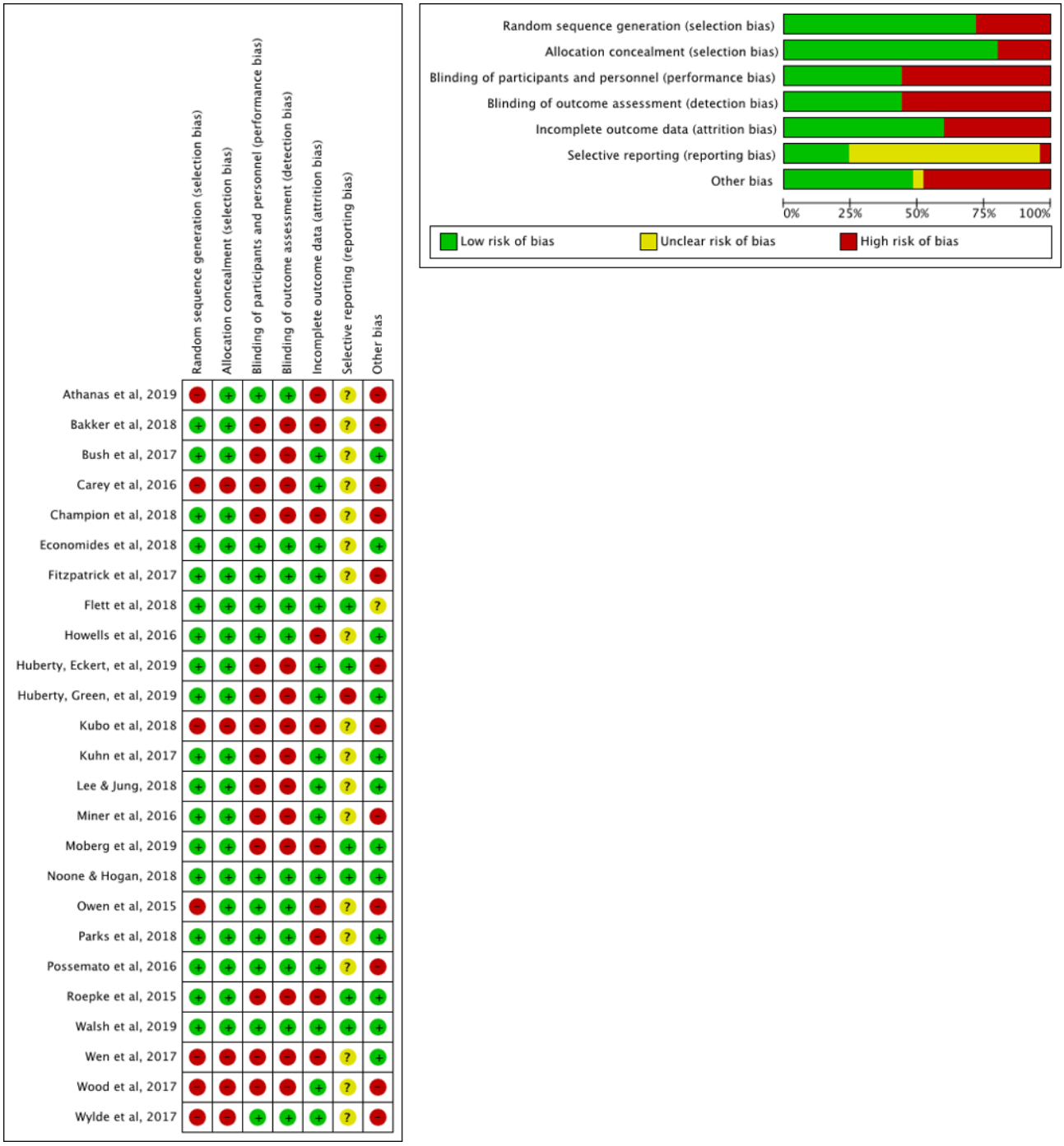


**Figure 4.** Summary of risk of bias.




## Discussion

### Principal Findings

To our knowledge, this is the first review of psychosocial wellness and stress management apps using a multilevel search strategy of mobile app search engines (ie, what self-help consumers would find) followed by a literature review (ie, what scientists would find). We aimed to explore treatment features and components commonly folded into mainstream apps and whether and how these differed from those of apps tested in research and clinical trials. In addition, we summarized the existing literature on all identified apps.

We identified 1009 stress management and psychosocial wellness apps on the Apple Store and Google Play. App contents and features were varied and eclectic, ranging from journaling to hypnosis. Of the 5 most common treatment features and components for all apps, only 1 was evidence based (mindfulness-meditation). Unsurprisingly, the subset of apps with research publications was much more cohesive; the 5 most common therapeutic components identified in this group of apps were all evidence based. We found that 2.08% (21/1009) of apps identified for inclusion had supporting research, and the majority of apps had only 1 research publication. All the published efficacy studies demonstrated some evidence that the app *works*, although effect sizes were rarely reported. However, the *file-drawer problem* [68] in academic research (ie, studies

with null findings are less likely to be published) and contrary business incentives (ie, publishing null findings does not make for a marketable app) may contribute to potential publication biases that highlight positive effects. The majority of published feasibility studies demonstrated some evidence of user satisfaction and acceptability. Our research expands on findings from classic methodology systematic reviews of smartphone-based anxiety and depression interventions that have found small to moderate positive effects [11,16].

Although there is a surfeit of mobile apps available for free download, only a small fraction of these have been tested in research settings. Even for the few published studies, the state of the science of mHealth for stress management and emotional wellness is in its nascent stages. Roughly half of the efficacy studies included in our review were either non-RCTs or RCTs without an active treatment comparison condition (ie, waitlist control). Approximately, one-fourth of all included studies were feasibility studies that did not measure efficacy outcomes. The majority of studies were not powered for analysis and, therefore, did not designate a priori primary vs secondary outcomes. Thus, the rapidly growing consumer market of mHealth for mental health is facing a similar *research-practice gap* to that of traditional face-to-face interventions: an overwhelming majority of self-help seekers may not be receiving evidence-based care [20]. This is not to discount the fact that businesses may have internally rigorous research and development processes outside of publications in scientific journals that we are unable to track or evaluate in a systematic way.

The World Health Organization (2019) recently released a guideline on digital health interventions for strengthening health systems based on an assessment of the benefits, harms, acceptability, feasibility, resource use, and equity considerations. The guideline's primary objective is the adoption of evidence-based interventions [69]. The European Commission provides complementary guidelines, including an assessment of data protection and privacy, safety, scientific content, and effectiveness [70]. In practice, mHealth interventions developed and tested in formal research settings for research purposes are rarely made available to the general public [71].

Potential future directions for traversing the research-practice divide are for academic researchers to partner with health technology companies and businesses to develop and test publicly available apps [72]. Such collaborations would improve the rigor of app development and continuous refinement by applying quality control standards to an unregulated market while capitalizing on the strengths of the commercial sector in financial and personnel resources, innovation, marketing, and motivational factors for user engagement. The app development process in the commercial sector adheres to a *user-centered design* framework, which engages end users as part of an iterative design process to better understand facilitators and barriers to sustainability and use [73]. This is a crucial model to apply in mHealth research because of historically low user adherence and retention rates. In addition, it is important to bolster the representation of other known evidence-based strategies in app components and features such as cognitive behavioral therapy, which was only represented in 3.07% (31/1009) apps identified. This was consistent with a previous review of popular anxiety and depression mobile apps that found evidence-based treatment strategies were poorly represented [74]. Finally, there is a need for comprehensive, consolidated, publicly available repositories of evidence-based stress management and psychosocial wellness apps going beyond consumer reports (including transparent information on public, private, or government ownership, public launch date, durability, and version history) so that the general public can make informed choices. Potential users should be directed to web-based resources such as PsyberGuide [75] to explore ratings and reviews for digital mental health products [71].

The majority of apps we identified were designed as self-help interventions; they were not necessarily intended for those with psychopathology. For individuals who are interested in seeking self-help via publicly available mHealth interventions, it is advisable to caution them against using this as a replacement to traditional treatment approaches, especially in the case of moderate to severe psychosocial problems. Importantly, the majority of studies found in our review used a sample of convenience (college students and users who have already downloaded the app and are therefore motivated to use them). It is unclear whether these apps would perform similarly if participants had more severe psychopathology symptoms.

Similarly, the role of health care providers and psychosocial clinicians in the mHealth space for an eclectic group of self-help seekers warrants exploration. E-therapist support is infrequently built into consumer apps (11/1009, 1.09%) of apps in our review). In addition to being resource intensive, this level of intervention may not be universally appealing or therapeutically indicated for generally healthy individuals interested in psychosocial self-care. On the other hand, in a clinical population with a serious mental health condition (PTSD), 1 of the studies included in our review found that clinician-supported app use outperformed self-directed app use [40]. This is consistent with previous literature regarding the therapist-patient relationship as a significant predictor of success in psychosocial treatment [76], and it remains to be determined whether app efficacy or engagement could be enhanced when paired with some form of clinician support. Future research should explore the optimal balance between clinician assistance and self-direction in mHealth and for which target audience (mental health support vs mental illness treatment support). At the very least, clinicians working with patients with diagnosed mental illnesses may choose to recommend specific evidence-based apps for skills practice and as a supplement to in-person therapy.

Recently, the Food and Drug Administration has released a timely policy report of its intent to provide regulatory oversight of mobile medical apps "for diagnosis of disease or other conditions, or the cure, mitigation, treatment, or prevention of a disease" and to "apply this oversight authority only to those software apps whose functionality could pose a risk to a patient's safety if the software apps were not to function as intended" [38]. Exceptions to oversight regulations are made for licensed professionals who create an app solely for use in their own practice or the manufacturing of mobile medical apps solely for use in research, teaching, and analysis and not for commercial use. This may influence the target audience for whom publicly available apps are developed as well as what is available for

free. As the majority of apps identified in this review were advertised to consumers under a broad wellness and self-care umbrella, there are important gaps in addressing the unique psychosocial needs of vulnerable groups such as individuals with psychological disorders, individuals with chronic illnesses, and youths. Another important area of research is to explore how users engage with digital health technologies including the leveraging of *big data* analytics; identifying factors that can enhance engagement and usability will help inform the design and optimization of apps for long-term appeal and sustainability [5]. These factors have not yet been explored.

### Limitations

There is no established *gold standard* for searching, evaluating, or reviewing digital health technologies. Recognizing that a methodology for using mobile app search engines to identify apps for review was nontraditional, we leveraged prior research on consumer apps to create a study-specific template and decision rules [17-19]. We relied on product pages on the Apple Store and Google Play to extract data on intervention characteristics, and specific search terms for our literature review which is not without its limitations. Although we provided transparency of our methods here, we recognized that they may not be reproducible. Similarly, it was impossible to construct a *stable, final* database of apps, given the quickly evolving mHealth landscape; new apps are developed, and old apps retired at a rapid rate. In the time frame in which our app store search was conducted, for example, 10.15% (114/1123) apps that were originally included in our database were no longer available 3 months later. Hence, our findings may lack the stability of classical systematic reviews. Similarly, it is possible that app names may have changed from the research design and testing phase reported in peer review publications to its official launch on the Apple Store or Google Play, or that other existing peer-reviewed publications may not have been identified utilizing our search strategy. For example, Happify's website lists two additional published efficacy studies that were not identified by our literature review search terms [77,78]. We were unable to conduct a meta-analysis of RCTs because of the differences in measures and measurement timepoints implemented across studies and relatively few trials with active comparison conditions.

Next, the apps we selected may not be representative of all mHealth programs. We included only free apps or free apps with in-app purchases. It is possible that paid apps significantly differ from free apps with respect to content and efficacy. Furthermore, some in-app purchases included access to an e-therapist, suggesting that secure services require greater monetary resources. As a result, access to effective digital health technologies may represent an unappreciated and important health disparity. However, it is unclear whether and the extent to which the involvement of an e-therapist bolsters outcomes and whether the benefits outweigh the costs; only 1.09% (11/1009) of apps included an e-therapist feature. In addition, the apps included in our review were limited to the English language. This may lead to a Western cultural bias in overrepresentation of apps with active coping strategies that reflect individualistic values and a personal sense of control over stressors [79-81].

### Conclusions

In merging traditional systematic review methodologies with a direct-to-consumer selection criteria for mHealth apps, our study findings suggest that few publicly available stress management and psychosocial wellness apps that are discoverable to self-help seekers are evidence based. Additional research is needed regarding the relative role and utility of mHealth in individual self-care and the care of persons with serious mental illness, the value of mHealth-clinician collaborations, access to mHealth for patients with different resources, and the relative durability of mHealth impact. Meanwhile, clinicians, investigators, and consumers may use findings from our systematic review to navigate the volume of emerging digital health interventions for stress management and wellness.

## Acknowledgments

NL is funded as an Implementation Science Scholar through the National Heart, Lung, and Blood Institute of the National Institutes of Health (grant number: 5K12 HL137940-02).

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies protocol.
[DOCX File , 16 KB - mhealth_v8i5e17798_app1.docx ]

Multimedia Appendix 2
Table of intervention components of n=21 stress management apps with published research.
[DOCX File , 26 KB - mhealth_v8i5e17798_app2.docx ]

Multimedia Appendix 3
Table of general information for n=21 stress management apps with published research.

[DOCX File , 22 KB - mhealth_v8i5e17798_app3.docx ]

Multimedia Appendix 4
Table of findings from efficacy studies (n=25).
[DOCX File , 51 KB - mhealth_v8i5e17798_app4.docx ]

Multimedia Appendix 5
Table of usability or feasibility studies (n=10).
[DOCX File , 26 KB - mhealth_v8i5e17798_app5.docx ]

## References


1. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 1;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
2. Anderson M, Jiang J. Pew Research Center. 2018. Teens, Social Media & Technology URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2020-04-02]
3. Pew Research Center. 2019. Mobile Fact Sheet URL: https://www.pewinternet.org/fact-sheet/mobile/, [accessed 2019-04-04]
4. Meeker M, Wu L. RMS Media: The Media Specialists. 2013. Internet Trends: D11 Conference URL: https://rmsmedia.co.za/wp-content/uploads/2013/08/Mary-Meeker-Internet-Trends-052913-final.pdf [accessed 2020-04-14]
5. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
6. Pohl M. Research2Guidance. 2017. 325,000 Mobile Health Apps Available in 2017-Android Now the Leading Mhealth Platform URL: https://research2guidance.com/ [accessed 2019-10-01]
7. Chib A, Lin SH. Theoretical advancements in mhealth: a systematic review of mobile apps. J Health Commun 2018;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
8. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
9. Bose J, Hedden S, Lipari R, Park-Lee E. SAMHSA: Substance Abuse and Mental Health Services. 2017. Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health URL: https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHNationalFindingsReport2018/NSDUHNationalFindingsReport2018.pdf [accessed 2019-08-01]
10. Anderson NB, Johnson SB, Belar C, Breckler S, Norda KC, Ballard D, et al. American Psychological Association. 2012. Stress in America: Our Health at Risk URL: https://www.apa.org/news/press/releases/stress/2011/final-2011.pdf [accessed 2019-10-01]
11. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
12. European Agency for Safety and Health at Work. 2014. Calculating the Cost of Work-related Stress And Psychosocial Risks URL: https://osha.europa.eu/en/publications/literature_reviews/calculating-the-cost-of-work-related-stress-and-psychosocial-risks [accessed 2019-07-01]
13. Anderson M. Pew Research Center. 2019. Mobile Technology and Home Broadband URL: https://www.pewinternet.org/2019/06/13/mobile-technology-and-home-broadband-2019/, [accessed 2019-06-19]
14. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the national comorbidity survey replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
15. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
16. 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;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
17. Weekly T, Walker N, Beck J, Akers S, Weaver M. A review of apps for calming, relaxation, and mindfulness interventions for pediatric palliative care patients. Children (Basel) 2018 Jan 26;5(2):16 [FREE Full text] [doi: 10.3390/children5020016] [Medline: 29373515]
18. Greenle MM, Morgan B, Sayani S, Meghani SH. Identifying mobile apps targeting palliative care patients and family members. J Palliat Med 2018 Oct;21(10):1380-1385. [doi: 10.1089/jpm.2018.0157] [Medline: 30312126]

19. Coulon SM, Monroe CM, West DS. A systematic, multi-domain review of mobile smartphone apps for evidence-based stress management. Am J Prev Med 2016 Jul;51(1):95-105. [doi: 10.1016/j.amepre.2016.01.026] [Medline: 26993534]
20. Livingston NA, Shingleton R, Heilman ME, Brief D. Self-help smartphone applications for alcohol use, PTSD, anxiety, and depression: addressing the new research-practice gap. J Technol Behav Sci 2019 May 9;4(2):139-151. [doi: 10.1007/s41347-019-00099-6]
21. Statista. 2020. Subscriber Share Held by Smartphone Operating Systems in the United States From 2012 to 2019 URL: https://www.statista.com/statistics/266572/market-share-held-by-smartphone-platforms-in-the-united-states/, [accessed 2019-07-01]
22. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 1;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
23. van Daele T, Hermans D, van Audenhove C, van den Bergh O. Stress reduction through psychoeducation: a meta- analytic review. Health Educ Behav 2012 Aug;39(4):474-485. [doi: 10.1177/1090198111419202] [Medline: 21986242]
24. Yates TM, Masten AS. Fostering the future: resilience theory and the practice of positive psychology. In: Linley PA, Joseph S, editors. Positive Psychology in Practice. New York City, USA: John Wiley & Sons; 2004:521-539.
25. Farquharson L, MacLeod AK. A brief goal-setting and planning intervention to improve well-being for people with psychiatric disorders. Psychother Psychosom 2014;83(2):122-124. [doi: 10.1159/000356332] [Medline: 24457329]
26. Gu J, Strauss C, Bond R, Cavanagh K. Corrigendum to 'how do mindfulness-based cognitive therapy and mindfulness-based stress reduction improve mental health and wellbeing? A systematic review and meta-analysis of mediation studies' [clinical psychology review 37 (2015) 1-12]. Clin Psychol Rev 2016 Nov;49:119. [doi: 10.1016/j.cpr.2016.09.011] [Medline: 27776684]
27. Compas BE, Jaser SS, Bettis AH, Watson KH, Gruhn MA, Dunbar JP, et al. Coping, emotion regulation, and psychopathology in childhood and adolescence: a meta-analysis and narrative review. Psychol Bull 2017 Sep;143(9):939-991. [doi: 10.1037/bul0000110] [Medline: 28616996]
28. Davis DE, Choe E, Meyers J, Wade N, Varjas K, Gifford A, et al. Thankful for the little things: a meta-analysis of gratitude interventions. J Couns Psychol 2016 Jan;63(1):20-31. [doi: 10.1037/cou0000107] [Medline: 26575348]
29. Alarcon GM, Bowling NA, Khazon S. Great expectations: a meta-analytic examination of optimism and hope. Pers Individ Differ 2013 May;54(7):821-827. [doi: 10.1016/j.paid.2012.12.004]
30. Layous K, Lyubomirsky S. The how, why, what, when, and who of happiness: mechanisms underlying the success of positive activity interventions. Positive Emot 2014:473-495. [doi: 10.1093/acprof:oso/9780199926725.003.0025]
31. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
32. Lieberman M. HubSpot Blog. 2014. 10 Stats About Inbound Marketing That Will Make Your Jaw Drop URL: https://blog.hubspot.com/insiders/inbound-marketing-stats [accessed 2019-11-01]
33. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The cochrane collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
34. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 6;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
35. Walsh KM, Saab BJ, Farb NA. Effects of a mindfulness meditation app on subjective well-being: active randomized controlled trial and experience sampling study. JMIR Ment Health 2019 Jan 8;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]
36. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. Psychiatr Serv 2017 Apr 1;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
37. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
38. Athanas AJ, McCorrison JM, Smalley S, Price J, Grady J, Campistron J, et al. Association between improvement in baseline mood and long-term use of a mindfulness and meditation app: observational study. JMIR Ment Health 2019 May 8;6(5):e12617 [FREE Full text] [doi: 10.2196/12617] [Medline: 31066704]
39. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
40. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]

41. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
42. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
43. Wood AE, Prins A, Bush NE, Hsia JF, Bourn LE, Earley MD, et al. Reduction of burnout in mental health care providers using the provider resilience mobile application. Community Ment Health J 2017 May;53(4):452-459. [doi: 10.1007/s10597-016-0076-5] [Medline: 28070775]
44. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 8;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
45. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
46. Flett JA, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
47. Haug NA. PsyberGuide. 2016. Headspace: An Expert Review URL: https://www.psyberguide.org/expert-review/headspace-expert-review/ [accessed 2019-07-01]
48. Kubo A, Altschuler A, Kurtovich E, Hendlish S, Laurent CA, Kolevska T, et al. A pilot mobile-based mindfulness intervention for cancer patients and their informal caregivers. Mindfulness (N Y) 2018 Dec;9(6):1885-1894 [FREE Full text] [doi: 10.1007/s12671-018-0931-2] [Medline: 30740187]
49. Wylde CM, Mahrer NE, Meyer RM, Gold JI. Mindfulness for novice pediatric nurses: smartphone application versus traditional intervention. J Pediatr Nurs 2017;36:205-212. [doi: 10.1016/j.pedn.2017.06.008] [Medline: 28888505]
50. Wen L, Sweeney TE, Welton L, Trockel M, Katznelson L. Encouraging mindfulness in medical house staff via smartphone app: a pilot study. Acad Psychiatry 2017 Oct;41(5):646-650. [doi: 10.1007/s40596-017-0768-3] [Medline: 28795335]
51. Noone C, Hogan MJ. A randomised active-controlled trial to examine the effects of an online mindfulness intervention on executive control, critical thinking and key thinking dispositions in a university student sample. BMC Psychol 2018 Apr 5;6(1):13 [FREE Full text] [doi: 10.1186/s40359-018-0226-3] [Medline: 29622047]
52. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
53. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
54. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
55. Parks AC, Williams AL, Tugade MM, Hokes KE, Honomichl RD, Zilca RD. Testing a scalable web and smartphone based intervention to improve depression, anxiety, and resilience: a randomized controlled trial. Int J Wellbeing 2018 Dec 8;8(2):22-67. [doi: 10.5502/ijw.v8i2.745]
56. Chittaro L, Vianello A. Evaluation of a mobile mindfulness app distributed through on-line stores: a 4-week study. Int J Hum-Comput St 2016 Feb;86:63-80. [doi: 10.1016/j.ijhcs.2015.09.004]
57. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app 'calm' to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
58. Lee RA, Jung ME. Evaluation of an mhealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
59. Huberty J, Eckert R, Larkey L, Kurka J, de Jesús SA, Yoo W, et al. Smartphone-based meditation for myeloproliferative neoplasm patients: feasibility study to inform future trials. JMIR Form Res 2019 Apr 29;3(2):e12662 [FREE Full text] [doi: 10.2196/12662] [Medline: 31033443]
60. Diefenbach S, Müssig A. Counterproductive effects of gamification: an analysis on the example of the gamified task manager Habitica. Int J Hum Comput St 2019 Jul;127:190-210. [doi: 10.1016/j.ijhcs.2018.09.004]
61. Linette D, Bryant J. Managing Stress in Health Care With Meditation: Got a Minute? Am Nurse Today 2017;12(10):- [FREE Full text]
62. Bush NE, Ouellette G, Kinn J. Utility of the T2 mood tracker mobile application among army warrior transition unit service members. Mil Med 2014 Dec;179(12):1453-1457. [doi: 10.7205/MILMED-D-14-00271] [Medline: 25469967]

63. Carey TA, Haviland J, Tai SJ, Vanags T, Mansell W. MindSurf: a pilot study to assess the usability and acceptability of a smartphone app designed to promote contentment, wellbeing, and goal achievement. BMC Psychiatry 2016 Dec 12;16(1):442 [FREE Full text] [doi: 10.1186/s12888-016-1168-z] [Medline: 27955643]
64. Griauzde D, Kullgren JT, Liestenfeltz B, Ansari T, Johnson EH, Fedewa A, et al. A mobile phone-based program to promote healthy behaviors among adults with prediabetes who declined participation in free diabetes prevention programs: mixed-methods pilot randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 9;7(1):e11267 [FREE Full text] [doi: 10.2196/11267] [Medline: 30626566]
65. Bakker D, Kazantzis N, Rickwood D, Rickard N. Development and pilot evaluation of smartphone-delivered cognitive behavior therapy strategies for mood- and anxiety-related problems: MoodMission. Cogn Behav Prac 2018 Nov;25(4):496-514. [doi: 10.1016/j.cbpra.2018.07.002]
66. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
67. Bush NE, Dobscha SK, Crumpton R, Denneson LM, Hoffman JE, Crain A, et al. A Virtual Hope Box smartphone app as an accessory to therapy: proof-of-concept in a clinical sample of veterans. Suicide Life Threat Behav 2015 Feb;45(1):1-9. [doi: 10.1111/sltb.12103] [Medline: 24828126]
68. Rosenthal R. The file drawer problem and tolerance for null results. Psychol Bull 1979;86(3):638-641. [doi: 10.1037/0033-2909.86.3.638]
69. World Health Organization. WHO Guideline Recommendations On Digital Interventions For Health System Strengthening. Geneva, Switzerland: World Health Organization; 2019.
70. European Commission. 2017. Report of the Working Group on mHealth Assessment Guidelines URL: https://ec.europa.eu/digital-single-market/en/news/report-working-group-mhealth-assessment-guidelines [accessed 2020-03-09]
71. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537. [doi: 10.1016/j.cbpra.2018.01.002]
72. Hingle M, Patrick H, Sacher PM, Sweet CC. The intersection of behavioral science and digital health: the case for academic-industry partnerships. Health Educ Behav 2019 Feb;46(1):5-9. [doi: 10.1177/1090198118788600] [Medline: 30041556]
73. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mhealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
74. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. Behav Res Ther 2019 Dec;123:103498. [doi: 10.1016/j.brat.2019.103498] [Medline: 31707224]
75. PsyberGuide. URL: https://psyberguide.org/ [accessed 2019-03-09]
76. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000 Jun;68(3):438-450. [doi: 10.1037/0022-006x.68.3.438] [Medline: 10883561]
77. Williams A, Parks A, Cormier G, Stafford J, Whillans A. Improving resilience among employees high in depression, anxiety, and workplace distress. Int J Manag Res 2018 Dec;9(1-2):4-22.
78. Carpenter J, Crutchley P, Zilca RD, Schwartz HA, Smith LK, Cobb AM, et al. Correction: Seeing the "Big" Picture: Big Data Methods for Exploring Relationships Between Usage, Language, and Outcome in Internet Intervention Data. J Med Internet Res 2017 Dec 19;19(12):e347 [FREE Full text] [doi: 10.2196/jmir.8099] [Medline: 29258059]
79. Kuo BC. Culture's consequences on coping. J Cross Cul Psychol 2010 Sep 9;42(6):1084-1100. [doi: 10.1177/0022022110381126]
80. Chun C, Moos R, Cronkite R. Culture: a fundamental context for the stress and coping paradigm. In: Handbook of Multicultural Perspectives on Stress and Coping. Boston, MA: Springer; 2006:29-53.
81. O'Connor DB, Shimizu M. Sense of personal control, stress and coping style: a cross-cultural study. Stress Health 2002 Oct;18(4):173-183. [doi: 10.1002/smi.939]


## Abbreviations

**e-therapist:** electronic therapist
**mHealth:** mobile health
**PTSD:** posttraumatic stress disorder
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by Q Yuan, M Stuckey, A Tariq; comments to author 03.03.20; revised version received 17.03.20; accepted 22.03.20; published 22.05.20.*

*Please cite as:*
*Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, Rosenberg AR*
*Android and iPhone Mobile Apps for Psychosocial Wellness and Stress Management: Systematic Search in App Stores and Literature Review*
*JMIR Mhealth Uhealth 2020;8(5):e17798*
*URL: http://mhealth.jmir.org/2020/5/e17798/*
*doi:10.2196/17798*
*PMID:32357125*

Original Paper

# Content and Quality of Infant Feeding Smartphone Apps: Five-Year Update on a Systematic Search and Evaluation

Heilok Cheng[1,2], BSc, MND; Alison Tutt[3*], DipHlthSc, BNurs, GradDipHlthSc (Midwifery), GradCert, MNurs; Catherine Llewellyn[3*], MNurs, PGCertMgt; Donna Size[3,4*], RN, CM, IBCLC, GradCert, GradDipAdvNurs, MAdvNurs; Jennifer Jones[1,3,5*], RN, RM, CFHN, ADCHN (CommNurs), GradDipAppSc(Com Hlth Nurs & Paed), MNurs; Sarah Taki[2,6,7], BSc, MND, PhD; Chris Rossiter[1], BA (Hons); Elizabeth Denney-Wilson[1,2,5], BNurs, MPH, PhD

[1]Susan Wakil School of Nursing and Midwifery, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia

[2]Centre of Research Excellence in Early Prevention of Obesity in Childhood, The University of Sydney, Camperdown, Australia

[3]Child and Family Health Nursing, Sydney Local Health District, NSW Health, Camperdown, Australia

[4]Lactation Clinic, Royal Prince Alfred Hospital, NSW Health, Camperdown, Australia

[5]Sydney Institute for Women, Children and their Families, Sydney Local Health District, NSW Health, Camperdown, Australia

[6]Health Promotion, Population Health Research & Evaluation Hub, Sydney Local Health District, NSW Health, Camperdown, Australia

[7]Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Camperdown, Australia

[*]these authors contributed equally

**Corresponding Author:**
Heilok Cheng, BSc, MND
Susan Wakil School of Nursing and Midwifery
Faculty of Medicine and Health
The University of Sydney
88 Mallett Street
Camperdown, 2050
Australia
Phone: 61 1800793864
Email: jessica.cheng@sydney.edu.au

## Abstract

**Background:** Parents use apps to access information on child health, but there are no standards for providing evidence-based advice, support, and information. Well-developed apps that promote appropriate infant feeding and play can support healthy growth and development. A 2015 systematic assessment of smartphone apps in Australia about infant feeding and play found that most apps had minimal information, with poor readability and app quality.

**Objective:** This study aimed to systematically evaluate the information and quality of smartphone apps providing information on breastfeeding, formula feeding, introducing solids, or infant play for consumers.

**Methods:** The Google Play store and Apple App Store were searched for free and paid Android and iPhone Operating System (iOS) apps using keywords for infant feeding, breastfeeding, formula feeding, and tummy time. The apps were evaluated between September 2018 and January 2019 for information content based on Australian guidelines, app quality using the 5-point Mobile App Rating Scale, readability, and suitability of health information.

**Results:** A total of 2196 unique apps were found and screened. Overall, 47 apps were evaluated, totaling 59 evaluations for apps across both the Android and iOS platforms. In all, 11 apps had affiliations to universities and health services as app developers, writers, or editors. Furthermore, 33 apps were commercially developed. The information contained within the apps was poor: 64% (38/59) of the evaluations found no or low coverage of information found in the Australian guidelines on infant feeding and activity, and 53% (31/59) of the evaluations found incomplete or incorrect information with regard to the depth of information provided. Subjective app assessment by health care practitioners on whether they would use, purchase, or recommend the app ranged from poor to acceptable (median 2.50). Objective assessment of the apps' engagement, functionality, aesthetics, and information was scored as acceptable (median 3.63). The median readability score for the apps was at the American Grade 8 reading level. The suitability of health information was rated superior or adequate for content, reading demand, layout, and interaction with the readers.

**Conclusions:** The quality of smartphone apps on infant feeding and activity was moderate based on the objective measurements of engagement, functionality, aesthetics, and information from a reliable source. The overall quality of information on infant feeding and activity was poor, indicated by low coverage of topics and incomplete or partially complete information. The key areas for improvement involved providing evidence-based information consistent with the Australian National Health and Medical Research Council's Infant Feeding Guidelines. Apps supported and developed by health care professionals with adequate health service funding can ensure that parents are provided with credible and reliable resources.

***(JMIR Mhealth Uhealth 2020;8(5):e17300)*** doi:10.2196/17300

**KEYWORDS**

breast feeding; bottle feeding; infant food; readability; consumer health information; breastfeeding; mobile apps; smartphones

## Introduction

### Background

Parents of a new baby can access a wealth of information and support from the internet through multiple electronic devices [1-3]. There is evidence to suggest, however, that web-based advice is not always evidence-based, even in the critical area of infant nutrition [4]. Smartphone ownership has expanded worldwide, with 81% of Australian adults and 97% of Australians aged 18 to 34 years owning a smartphone in 2018 [5]. Recent national data suggest that 46% of Australian adults accessed the internet for health information [6]. In 2017, 84% of Australian adults used mobile phones to access the internet, exceeding access through laptop computers (69%) and desktop computers (54%) [7]. Furthermore, smartphone ownership has now surpassed the ownership of desktop and laptop computers [8].

The proliferation of web-based health information sources is reflected by the growing literature for health care professionals discussing and advising the use of new technology [9-12]. Studies have shown that parents and pregnant women trusted hospital, government, and university websites as accurate, regulated, useful, and current sources of pregnancy and parenting information [13,14]. Parents in a video education study on introducing solid foods preferred the internet as a source of information for infant nutrition and felt that public authorities were important information providers [15].

Governments and nonprofit organizations have developed smartphone apps to promote and enhance breastfeeding [16-20]. A recent content analysis of social support in 31 breastfeeding apps found that the most common topics were managing breastfeeding problems (informational support) and locating where to express or breastfeed in public (instrumental support) [21]. Increasingly, clinical trials have assessed mobile health (mHealth) and smartphone apps as interventions to promote and support breastfeeding among mothers and their partners [22-24] through text messaging, goal setting, access to information and videos, provision of online support groups, and troubleshooting breastfeeding difficulties. A review by Tang et al [25] on digital interventions that support breastfeeding found that client communication systems to communicate breastfeeding information, facilitate communication, and provide on-demand information services through text messages, phone calls, email, smartphone apps, and websites may improve breastfeeding adherence.

In Australia, government health services use mHealth apps to support routine child and family health nursing practice in fields such as child literacy and development [26,27], immunization [28], and safe infant sleeping [29]. An early childhood obesity prevention trial will test an app developed for parents and caregivers [30] to be integrated into an Australian statewide pregnancy coaching service [31]. Clearly, apps used as part of routine care must meet practice standards for providing understandable, reliable, current, and evidence-based information independent of commercial associations. This highlights the need for evidence-based, well-developed, and updated apps.

Despite the proliferation of apps and their increasing popularity with parents, their quality may not have kept pace with their quantity. Earlier research on infant feeding apps and websites available in Australia found that 78% of the apps were of poor quality, with deficits in the breadth and completeness of information, author credibility, and readability [4]. A similar review of infant feeding apps in China found that most apps advertised infant formula and parenting products and rated poorly on the availability of information, author credibility, and transparency in disclosing advertising policy, app ownership, and app sponsorship [32]. A recent review of mHealth apps for parents of infants in neonatal intensive care units found that smartphone apps were functional but had low quality and credibility, with only 2 apps rated *good* by nurse and information scientist reviewers and only 5 apps deemed *trustworthy* [33].

Since the 2015 study [4], the Australian ownership of smartphones has increased from 64% to 81%, and the proportion of users accessing the internet through mobile phones has increased from 42% to 84% [7,34]. In 2017, there were 325,000 health, fitness, and medical apps available [35]. The increasing interest in infant feeding and physical activity apps indicated by popular search queries in the Google search engine (Multimedia Appendix 1) demonstrates the need to systematically assess and update the current smartphone app landscape. Furthermore, there is a continual turnover of smartphone apps, with several apps from the 2015 study [4] subsequently removed from distribution.

### Aim

Given the rapid expansion of digital technology into the realm of child health, this study aimed to evaluate the quality of information on infant nutrition and physical activity currently available to Australian parents via smartphone apps. It updates and expands the 2015 systematic assessment, examining the

comprehensibility, suitability, and readability of information in free and paid apps available in Australia.

## Methods

### Study Design

This study used systematic methods to identify, select, and evaluate infant feeding and activity apps that were available in Australia between August 2018 and January 2019. It assessed aspects of their quality and utility using validated and purpose-specific instruments to replicate and update an earlier study [4]. Full details of the methods are given in Multimedia Appendix 2, and evaluation tools are described in Multimedia Appendix 3.

#### *Stage 1: App Selection*

Smartphone apps were identified by searching app platforms from the two largest smartphone operating systems: App Store for iPhone Operating System (iOS; Apple Inc) and Google Play for Android (Google LLC). The search terms included variations in *infant feeding, baby feeding, breastfeeding, formula feeding, bottle feeding, baby food, baby weaning, infant activity,* and *tummy time.*

We used the search engine Google Play on a desktop for searching Android apps [36]. It was not possible to search the App Store on a desktop [37]; therefore, we conducted all App Store searches on the authors' iOS smartphones.

Members of the research team screened all located apps for eligibility: 4 authors screened iOS apps, and 2 authors screened Android apps. The first author cross-checked all apps. Apps were reviewed if they met the inclusion criteria. Any disagreements regarding the inclusion of apps in the study were discussed until consensus was reached.

The inclusion criteria for selection included apps written in English, targeted at parents of infants up to one year of age, and containing information on at least one of the following topics: milk feeding behaviors (breastfeeding, formula feeding, expressing breast milk, frequency or timing of feeding, and correct preparation of infant formula), solid food feeding behaviors (age of introduction, types of food introduced, repeated exposure, varied exposure, and reducing exposure to unhealthy food and beverages), or infant activity (tummy time, infant play, and movement).

In the selection stage, we excluded apps that were inaccessible with dead or broken links; were formatted as electronic books, news, magazines, podcasts, blogs, or word documents; had restricted access; did not have an English language option or were machine-translated into English; were games or gaming apps, or contained stolen or farmed content [38] from other apps or websites. Examples of excluded apps are given in Multimedia Appendix 2. Apps whose main function was to monitor or time infant care tasks, without providing any educational information on infant feeding and activity, were also excluded.

#### *Stage 2: App Evaluation*

In this stage, reviewers evaluated the selected apps on several dimensions (coverage and depth of information, quality, data security, and app accessibility, suitability, and readability) using a range of instruments. The reviewers rated all instruments using the Research Electronic Data Capture platform (Vanderbilt University) [39].

#### Coverage and Depth of Information

Accurate coverage (referred to hereafter as *coverage*) and depth of information were evaluated using a quantitative tool developed for the 2015 study [4], based on the Australian government's guidelines on infant feeding [40] and physical activity [41], with permission from the tool's authors. It has 9 topics with 22 subtopics on encouraging and supporting breastfeeding, initiating breastfeeding, establishing and maintaining breastfeeding, managing common breastfeeding problems, expressing and storing breast milk, breastfeeding in specific situations, preparing and using infant formula, introducing solid foods, and encouraging infant activity.

Coverage, defined as the breadth of the subtopics correctly covered in each app, was scored as either correct (+1), incorrect (−1), not addressed (0), or not applicable. Depth, defined as the completeness of information covered in each subtopic, was scored as partially complete (+0.5), complete (+1), or incompletely addressed or incorrect information (0). If the subtopic was scored incorrectly for coverage, it was automatically scored as incorrect for depth.

We also rated each app's coverage using the criteria from the Health-Related Website Evaluation Form [42]. Coverage was summarized as excellent (≥90%), adequate (75%-89%), or poor (≤74%). Depth was summarized as complete (100%), partial (50%-99%), or low or no (≤49%) completeness.

#### App Quality

App quality was evaluated using the Mobile App Rating Scale (MARS) [43], which was not available at the time of the original study in 2015. The MARS is a 23-item quality rating tool that uses a 5-point rating scale, scored as 1 (inadequate), 2 (poor), 3 (acceptable), 4 (good), and 5 (excellent), with 4 objective scales on engagement (5 domains on interesting, fun, or interactive content), functionality (4 domains on app navigation and logical usability), aesthetics (3 domains on graphic design and visual appeal), and information quality (7 domains on credibility of source). The MARS also includes 1 subjective quality scale incorporating the user's judgment on their likelihood of recommending, using, and purchasing the app and a personal 5-star rating. This is reported separately as the *subjective MARS score*.

A final measurement of app quality, the objective MARS score, is calculated as a 5-star rating using the mean from the scores from the objective (engagement, functionality, aesthetics, and information quality) scales.

#### App Usability

The authors of the previous study [4] developed 2 additional scales not included in other app assessment tools available or found by the authors during the app evaluation stage. These were data security, with items assessing data encryption and privacy, and accessibility, incorporating multilanguage options, one-handed functionality, and availability of help guides. We

added these scales to the MARS tool to create the *modified MARS score*, which we calculated separately from the objective MARS score.

### Suitability of Information

The appropriateness of the information on the apps was evaluated using the Suitability Assessment of Materials (SAM) tool [44]. The SAM is a 22-item validated instrument that assesses content, literacy level, graphics, layout, interaction with readers, and cultural appropriateness. Each item is scored as superior (+2), adequate (+1), not suitable (0), or not applicable. The sum of the scores of the items generates a final score summarized as superior (70%-100%), adequate (40%-69%), or not suitable (0%-39%) appropriateness of information for the target audience.

The hypothetical target audience used in the app evaluation was Australians with a year 3 to 4 reading level, with or without a multicultural background [45].

### Readability

The SAM also assesses the readability, or grade level of written text, measured using the Flesch-Kincaid (F-K) [46] and the Simple Measure of Gobbledygook (SMOG) [47] tools. The reviewers assessed the readability by typing a section of writing from each app into an online readability calculator [48] that calculated F-K and SMOG scores; they also used Microsoft Word software (2010 and later; Microsoft Corporation) [49] to generate an alternative F-K score. Each reviewer selected the passage of the text they assessed.

The F-K and SMOG scores are reported as American reading grades. The Australian federal government's Plain English guidelines [45] recommend writing for a reading level of Australian school year 3 to 4, the equivalent of American Grade 3 to 4 reading level. The South Australian state government's health literacy guidelines recommend writing for a reading level of Australian year 8 [50].

Readability is an item in the SAM. Using the SAM, we summarized F-K and SMOG readability scores as superior (Grade 5 and lower reading level), adequate (Grade 6 to 8 reading level), or not suitable (Grade 9 or higher reading level).

### Interrater Reliability

We undertook interrater reliability (IRR) testing with 2 or more reviewers assessing at least 20% of the selected apps using all rating tools. We tested apps that were available on both iOS and Android platforms. Discrepancies were discussed until reviewers reached a consensus on their final ratings.

IRR was calculated for the readability scores, MARS scores, SAM, and the evaluation of information content using Krippendorff α (α), which is appropriate when there are missing or incomplete data. Using Krippendorff standards for data reliability, .667≤α<.80 was accepted as tentatively reliable, and α≥.80 was accepted as reliable.

### Statistics

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 25.0 (IBM Corporation). Descriptive results on the ratings from the various instruments are reported as median (IQR). We calculated the correlations among different measures of readability using the Spearman rank-order correlation. We also compared the 5-point rating scores in the MARS tool with the user ratings of the apps presented in the Apple App Store and Google Play Store using the Pearson correlation coefficient. Krippendorff α was calculated using the KALPHA macro for IBM SPSS [51]. Significant values were indicated at $P<.05$.

### Ethics Approval

This study did not require ethics approval.

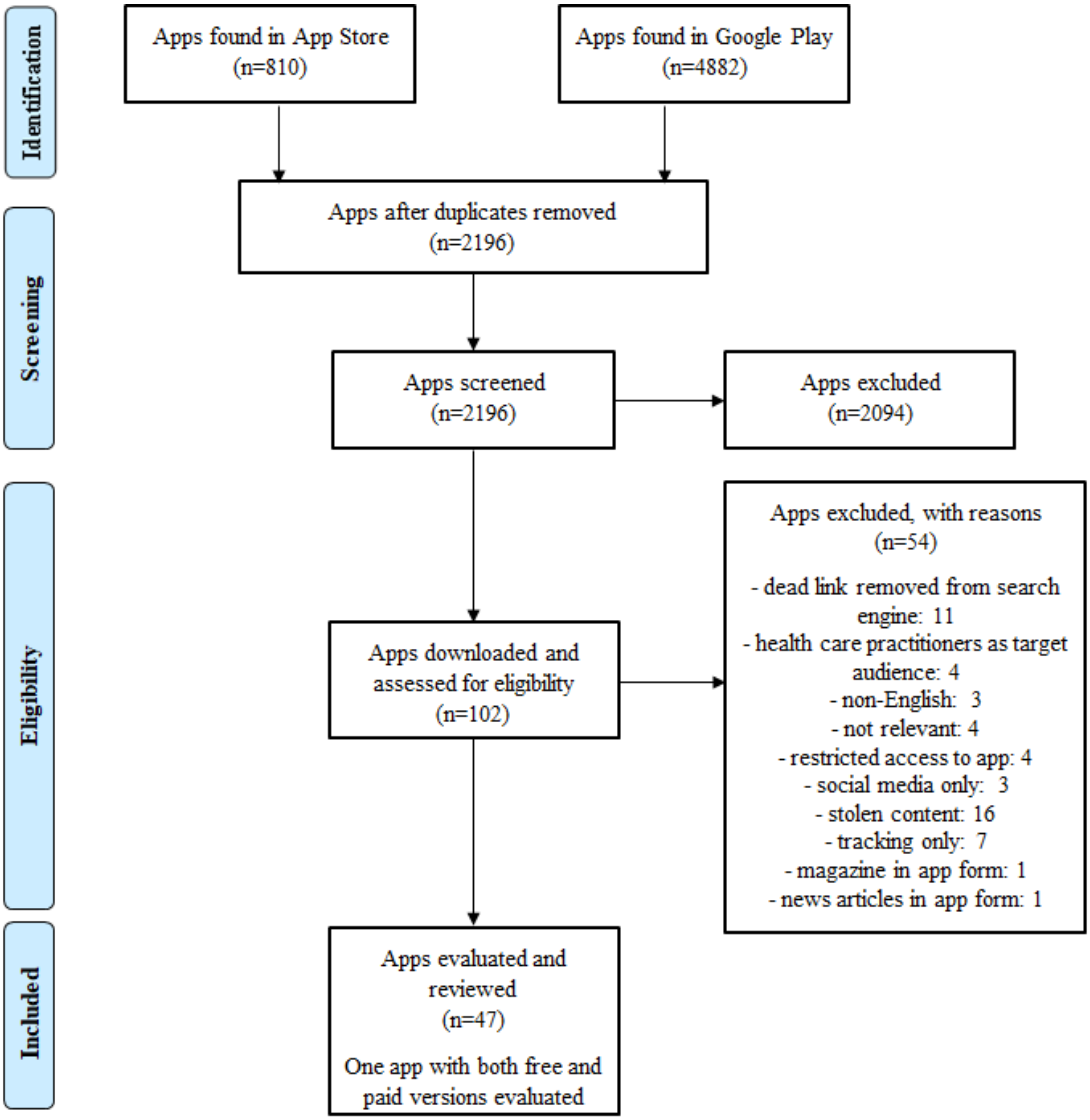
## Results

### Stage 1: Smartphone App Selection

#### *Screening Process*

App searches were performed between August and September 2018. A total of 5692 apps were identified for screening (Figure 1), with 3496 duplicates removed and 2196 apps screened for potential inclusion. After screening, 102 apps were downloaded for evaluation. Of these, 54 were excluded, and 47 were reviewed between September 2018 and January 2019. The apps included in this study are described in Multimedia Appendix 4. We undertook 59 evaluations for 47 apps, including 27 and 32 evaluations on iOS and Android smartphones, respectively.

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram of smartphone app selection process.




### *Description of the Selected Apps*

Overall, 1 app was developed with nationally competitive research funding, 10 apps were developed by the government or had government affiliations, 3 apps were developed by universities or had university affiliations, and 33 were commercial.

Furthermore, 2 apps were trialed with surveys [52,53]. One app was used in a randomized controlled trial but was not objectively evaluated [54]. Of the 47 apps, 35 apps (74%) were free to access, although 12 of these required payment to remove advertisements; access additional content, functions, and information; or access the full app without preview restrictions. Overall, 12 of the 47 apps (26%) were only accessible by purchase. A total of 10 apps were, as reported by app developers, available in languages other than English (eg, Arabic, Bosnian, Chinese, Danish, Dutch, French, German, Hindi, Italian, Japanese, Polish, Portuguese, Russian, Spanish, Swedish, and Vietnamese).

Of the 47 apps, 32 apps (68%) were located through infant feeding–related search terms, 13 (28%) through terms related to introduction to solids, and 25 (53%) through infant activity–related terms. Several apps were found through multiple keyword search terms, for example, in search terms related to infant feeding and introduction to solids.

## Stage 2: Smartphone App Evaluation

### *Interrater Reliability*

Overall, 12 of the 47 apps (26%) were evaluated independently by 2 different reviewers using iOS and Android platforms (Multimedia Appendix 4) to assess the IRR.

There was reliable IRR agreement ($\alpha \geq .80$) for the objective MARS ratings and the depth of information content in the infant activity subtopic. There was acceptable IRR agreement ($.667 \leq \alpha < .80$) for the modified MARS ratings, depth of information content in the infant feeding subtopic, and coverage of information content in the infant activity subtopic.

Although the IRR scores were relatively low on several instruments, the reviewers discussed discrepancies and the interpretation of evaluation criteria to ensure greater unanimity in future scoring. The following tables present results for 59 evaluations, taking into account the 12 apps rated by the 2 reviewers.

### *Coverage and Depth of Information*

The coverage and depth of information provided by most apps were relatively poor (Table 1). Assessment with the content evaluation tool (Multimedia Appendix 3) found a median coverage, or breadth of subtopics correctly covered, of 64% (IQR 40%-87%; Figure 2). The depth, or completeness of information covered in the subtopics, had a median rating of 48% (IQR 32%-67%). The majority of app evaluations (31/59, 53%) had *low or no completeness* of subtopics (Tables 2 and 3). Furthermore, two-thirds of the app evaluations (38/59, 64%) showed that the apps had *poor* depth of information, and only 7% (4/59) of the app evaluations rated information as *complete* (Tables 2 and 3).

Detailed reporting of the coverage and depth of information within each subtopic are shown in Multimedia Appendix 4. Subtopics pertinent to clinicians are described below under *Coverage of Information in Subtopics* and *Depth of Information in Subtopics* and in Table 4.

**Table 1.** The quantitative coverage and depth of information based on Australian infant feeding and physical activity guidelines in all apps (N=59 evaluations of 47 apps).

| Information quality | Median (%) | IQR (%) | Range (%) | Number of evaluations, n |
|---|---|---|---|---|
| **Coverage** | | | | |
| All apps | 64 | 40-87 | −20 to 100[a] | 59 |
| Infant feeding apps | 67 | 40-80 | 7 to 100 | 37 |
| Introduction to solid foods apps | 50 | 0-88 | −100 to 100[a] | 37 |
| Infant activity apps | 100 | 50-100 | 0 to 100 | 33 |
| **Depth** | | | | |
| All apps | 48 | 32-67 | 4 to 100 | 59 |
| Infant feeding apps | 38 | 21-50 | 0 to 86 | 37 |
| Introduction to solid foods apps | 38 | 6-50 | 0 to 100 | 37 |
| Infant activity apps | 50 | 50-88 | 0 to 100 | 33 |

[a]Results with negative scores indicate apps with overall negative scoring for topics reported incorrectly.

**Figure 2.** Quantitative evaluation of the coverage of information and the depth of information across all subtopics on infant feeding, introduction to solids, and physical activity in apps. A total of 59 evaluations were conducted for 47 apps. Median and IQR reported.

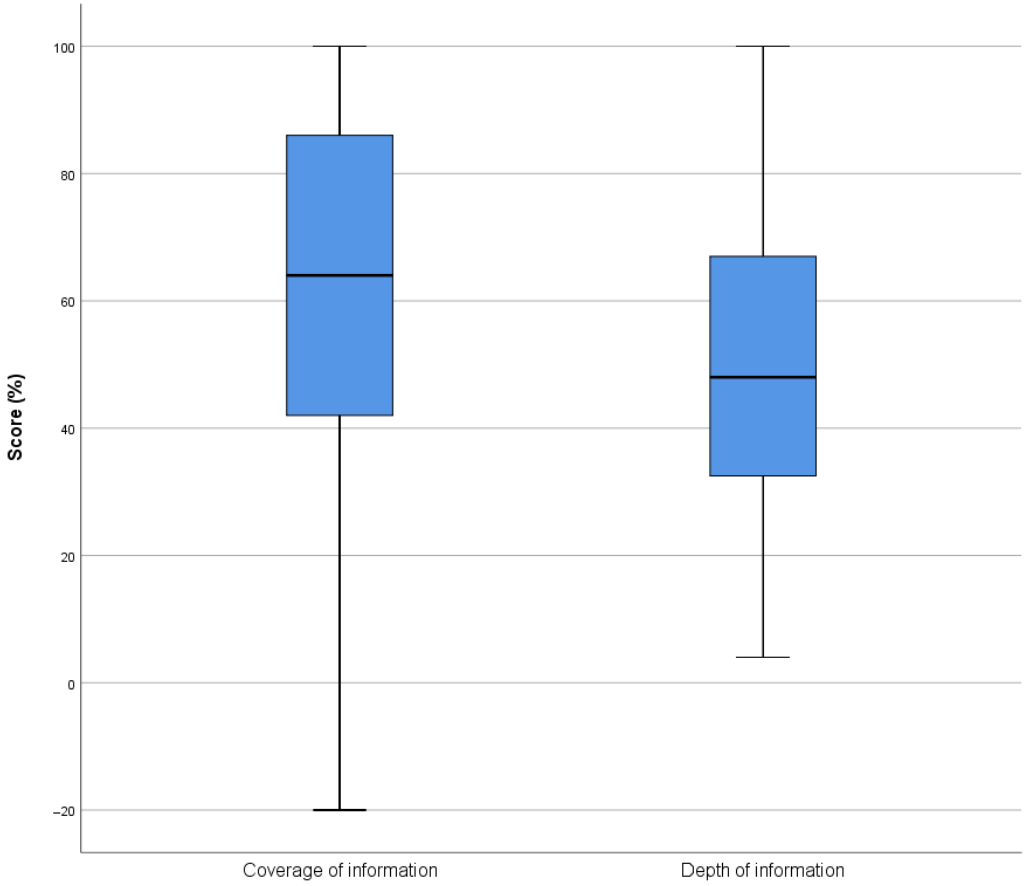



**Table 2.** The qualitative evaluation of coverage of information based on Australian infant feeding and physical activity guidelines in all apps (N=59 evaluations of 47 apps).

| Coverage of information | Poor, n (%) | Adequate, n (%) | Excellent, n (%) | Number of evaluations, n |
|---|---|---|---|---|
| All apps | 38 (64) | 8 (14) | 13 (22) | 59 |
| Infant feeding apps | 25 (68) | 6 (16) | 6 (16) | 37 |
| Introduction to solid foods apps | 26 (70) | 2 (5) | 9 (24) | 37 |
| Infant activity apps | 10 (29) | 0 (0) | 24 (71) | 33 |

**Table 3.** The qualitative evaluation of depth of information based on Australian infant feeding and physical activity guidelines in all apps (N=59 evaluations of 47 apps).

| Depth of information | Low or no completeness, n (%) | Partial completeness, n (%) | Complete, n (%) | Number of evaluations, n |
|---|---|---|---|---|
| All apps | 31 (53) | 24 (41) | 4 (7) | 59 |
| Infant feeding apps | 27 (73) | 10 (27) | 0 (0) | 37 |
| Introduction to solid foods apps | 26 (70) | 9 (24) | 2 (5) | 37 |
| Infant activity apps | 7 (21) | 8 (56) | 19 (24) | 33 |

**Table 4.** The depth (completeness) of information in the subtopics reported (N=59 evaluations of 47 apps).

| Completeness of information | Complete, n | Partially complete, n | Incorrect or incomplete, n | Number of evaluations, n[a] |
|---|---|---|---|---|
| **Encouraging and supporting breastfeeding** | | | | |
| Breastfeeding as the physiological norm | 16 | 11 | 5 | 32 |
| Protection and promotion of breastfeeding | 6 | 14 | 1 | 21 |
| Breastfeeding education for parents | 12 | 12 | 1 | 25 |
| **Initiating breastfeeding** | | | | |
| Physiology of breast milk and breastfeeding | 4 | 17 | 1 | 22 |
| The first breastfeed | 3 | 6 | 5 | 14 |
| **Establishing and maintaining breastfeeding** | | | | |
| Difficulty establishing breastfeeding | 1 | 10 | 2 | 13 |
| Factors affecting establishment of breastfeeding | 1 | 10 | 5 | 16 |
| Monitoring an infant's progress | 8 | 14 | 4 | 26 |
| Maternal nutrition | 4 | 11 | 7 | 22 |
| **Breastfeeding, common problems, and their management** | | | | |
| Maternal factors affecting breastfeeding | 3 | 15 | 1 | 19 |
| Infant factors affecting breastfeeding | 1 | 12 | 3 | 16 |
| **Expressing and storing breast milk** | | | | |
| Expressing breast milk | 5 | 13 | 5 | 23 |
| Feeding with expressed breast milk | 5 | 5 | 1 | 11 |
| Storage of expressed breast milk | 10 | 5 | 9 | 24 |
| **Breastfeeding in special situations** | | | | |
| Tobacco, alcohol, and other drugs | 2 | 13 | 6 | 21 |
| **Infant formula** | | | | |
| Preparing infant formula | 2 | 4 | 7 | 13 |
| Using infant formula | 6 | 6 | 4 | 16 |
| Special infant formula | 1 | 5 | 0 | 6 |
| **Introducing solids** | | | | |
| When should solid foods be introduced? | 7 | 15 | 13 | 35 |
| What foods should be introduced? | 3 | 3 | 21 | 27 |
| Foods and beverages most suitable for infants or foods that should be used in care | 4 | 14 | 9 | 27 |
| Healthy foods in the first 12 months (continued exposure and opportunity to sample a wide variety of healthy foods) | 10 | 9 | 5 | 24 |
| **Infant activity** | | | | |
| Encouraging physical activity for infants from birth for healthy development | 16 | 11 | 1 | 28 |
| Advice on types of infant physical activity and movements for development, including reaching and grasping; pulling and pushing; moving their head, body, and limbs during daily routines; and supervised floor play, including tummy time | 10 | 13 | 2 | 25 |

[a]Not all apps included information on all subtopics.

### *Coverage of Information in Subtopics*

Information coverage was lowest in apps related to infant feeding and introduction to solid foods. Infant activity subtopics were more likely to provide correct advice, with only 1 app reporting incorrect advice on encouraging infant activity for healthy development.

For information on infant feeding, incorrect advice was most frequently reported in the following subtopics: maternal nutrition during breastfeeding (3 apps), expressing and storing breast milk (3 apps across topics), breastfeeding in specific situations (tobacco, alcohol, and other drug use; 3 apps), and preparing and using infant formula (2-4 apps).

For information on introducing solid foods, incorrect advice was most frequently reported on the time to introduce solid foods (9 apps), first foods to introduce (15 apps), foods and beverages most suitable for infants (5 apps), and exposure to healthy foods for the first 12 months (3 apps).

### *Depth of Information in Subtopics*

The ratings on the depth of information were lowest in apps related to infant feeding and introduction to solid foods (Table 4).

In apps on infant feeding, the depth of information was best reported for special infant formula, with 3 apps reporting *partially complete* information and 1 app reporting *complete* information. *Incomplete or incorrect* information was reported in all other subtopics across infant feeding apps. *Partially complete* information on infant feeding was more frequently reported than *complete* information.

In apps on introduction to solid foods, *incomplete or incorrect* information was reported in all subtopics. For appropriate first foods, *incomplete or incorrect* information was reported more frequently than *correct* information. *Partially complete* information on when to introduce solid foods and foods and beverages most suitable for infants was more frequently reported than *complete* information.

Most apps on infant activity reported *partially complete* information on encouraging infant activity for healthy development and *complete* information on the types of different infant physical activities or movements for development.

### *App Quality*

Table 5 and Figure 3 present the results of app quality evaluation using the MARS tool, which rates different objective and subjective dimensions of app quality.

**Table 5.** Mobile App Rating Scale quality ratings (N=59 evaluations of 47 apps).

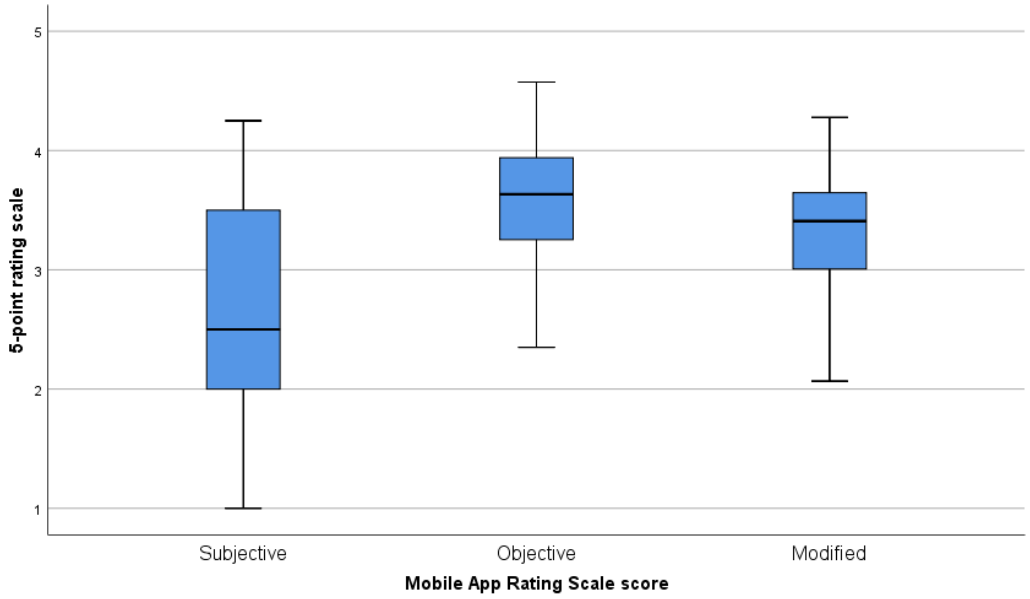
| MARS[a] evaluation scores | Median | IQR | Range | Apps rated good, n | Apps rated excellent, n |
|---|---|---|---|---|---|
| Engagement subscale | 3.00 | 2.60-3.40 | 1.80-4.20 | 12 | 0 |
| Functionality subscale | 4.25 | 3.75-4.75 | 2.0-5.0 | 26 | 24 |
| Aesthetics subscale | 4.33 | 4.0-4.67 | 2.0-5.0 | 32 | 10 |
| Information quality subscale | 3.60 | 3.0-3.80 | 1.75 -4.8 | 30 | 1 |
| Objective MARS score[b] | 3.63 | 3.24-3.99 | 2.07-4.28 | 38 | 1 |
| Data security subscale | 2.33 | 1.0-3.33 | 1.0-5.0 | 5 | 3 |
| Accessibility subscale | 3.00 | 2.0-3.67 | 1.0-5.0 | 18 | 1 |
| Modified MARS score[c] | 3.41 | 2.99-3.64 | 2.35-4.57 | 25 | 0 |
| Subjective MARS score | 2.50 | 2.0-3.5 | 1.0-4.25 | 20 | 1 |

[a]MARS: Mobile App Rating Scale.

[b]Objective MARS score=mean of engagement subscale+functionality subscale+aesthetics subscale+information quality subscale.

[c]Modified MARS score=mean of objective MARS score+data security subscale+accessibility subscale.

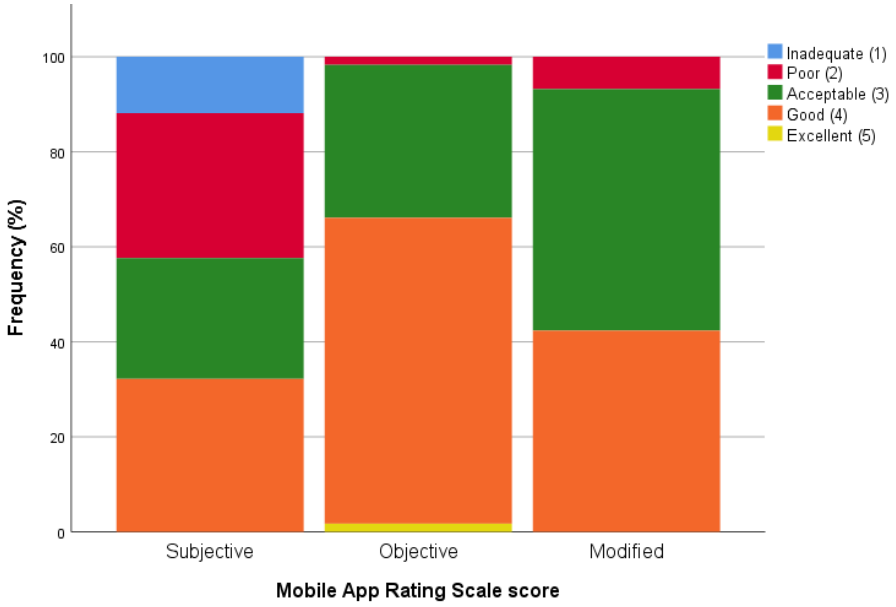
**Figure 3.** Quantitative evaluation of the 5-point Mobile App Rating Scale scores.




Overall, the quality of the apps was found to be mixed. Ratings were typically higher for items included in the objective score (especially the aesthetics and functionality subscales) than the subjective score. The inclusion of the 2 items on data security and accessibility lowered the median quality ratings of the apps, especially given that most apps rated poorly on data security (median 2.33; IQR 1.00-3.33).

Figure 4 indicates that very few apps were rated *excellent* across all items on the MARS scales, although higher proportions were rated *good*, especially for the objective subscales.

**Figure 4.** Qualitative evaluation of the 5-point Mobile App Rating Scale scores.




On the subjective MARS scale, out of 47 apps, there were 21 apps (45%) that the reviewer reported they would never recommend to others or recommend to very few people. Furthermore, reviewers, who were child health clinicians or health researchers, reported that they would not pay to access 30 of the 47 apps (64%), although 4 of these were free, developed by the government or a community health organization.

A total of 31 apps reported the authors' qualifications with health expertise, including doctors, nurses, lactation consultants, midwives, psychologists, physiotherapists, physical therapists, speech language therapists, health promotion officers, dietitians, nutritionists, occupational therapists, and sport therapists. As noted, 33 apps were developed by commercial entities, including 15 owned or developed by health care practitioners and 9 that consulted with health practitioners.

More detailed findings on the MARS and scores are reported in Multimedia Appendix 4.

We compared the MARS quality scores with the user-rated scores from the Apple App Store and Google Play Store reported in January 2019 (Multimedia Appendix 4). There was no significant correlation between the users' app ratings and any of the MARS scores, but there was a significant correlation

among the subjective, modified, and objective MARS scores ($P$<.001 each, 2-tailed).

### *Suitability of Infomation*

Overall, 42% apps (25/59 evaluations) were rated *superior* for suitability of health information, 54% apps (32/59 evaluations) were rated as *adequate*, and 3% apps (2/59 evaluations) were *not suitable*. More detailed findings on the SAM scores are reported in Multimedia Appendix 4.

Few apps were rated *superior* on the cultural appropriateness items, although most apps were considered *adequate* on the cultural match for an Australian setting. A few apps were considered *superior* when they were suitable for an Australian setting and featured information for a non-Western culture and demography. Few apps presented information with representation of images and examples demonstrating cultural diversity. The 3 apps that contained culturally diverse images were all developed outside of Australia (but were available in English); the target populations were Croatian (*Baby Food Chart*), mainland Chinese and Hong Kong Chinese (*Info for Nursing Mum*), and Maori and Pacific Islander families (*Raising Children*).

Many apps provided instructions for taking clear and specific actions, with topics subdivided to motivate users—for example, in apps on introducing solid foods, information was subdivided into sections on the types of food to introduce, how to prepare food, how to feed infants, and how to encourage dietary variety. Most information was provided in a question-and-answer format, which was rated as *adequate* reader interaction.

### *Readability*

The reading grade of app content was consistent across the tools used (Table 6).

There was a good correlation among reading grade scores using the 3 readability measures ($P$<.001; Multimedia Appendix 4).

**Table 6.** Readability scores of infant feeding and activity apps.

| American grade level reading score | Median | IQR |
|---|---|---|
| Flesch-Kincaid score (online tool) | 8 | 6-10 |
| Flesch-Kincaid score (Microsoft Word) | 8 | 6-9 |
| Simple Measure of Gobbledygook score | 8 | 6-9 |

However, very few apps met the Australian federal government's recommended level for written health information of Grade 4 reading level or below: either 3 apps (using the F-K online tool), 4 apps (F-K in Word), or 5 apps (SMOG). A majority of apps met the South Australian government's recommended level of Grade 8 level reading and below: 24 and 32 apps using F-K online tool and Microsoft Word tool, respectively, and 34 apps using the SMOG tool.

There was a low correlation among reading grades rated as not suitable, adequate, or superior in the SAM tool and the overall SAM score on adequacy of health information ($r$=0.25; $P$=.06).

### *Comparison With the 2015 Study*

The 2015 study [4] used similar methods to evaluate infant feeding apps for parents. Although this study aimed to replicate it, some of the methods of analysis and presentation differed. Table 7 presents comparable findings between the 2 studies.

**Table 7.** Comparison of evaluation outcomes used in the original 2015 study and this study.

| Instrument | This study (47 apps and 59 evaluations) | 2015 study (46 apps and 46 evaluations) |
|---|---|---|
| **Content evaluation tool (%), median (IQR)** | | |
| Coverage of information | 64 (40-87) | 65 (58-71) |
| Depth of information | 48 (32-67) | Reported graphically |
| **App quality using Quality Component Scoring System (scored out of 100%)** | | |
| Median (IQR) | Not undertaken | 49 (41-60) |
| Proportion rated poor (<50% score) | Not undertaken | 91 |
| **App quality using Mobile App Rating Scale (scored out of 5 points)** | | |
| **Objective scale** | | |
| Median (IQR) | 3.63 (3.24-3.99) | Not developed at the time of writing |
| Proportion rated poor (%, ≤2.5 score) | 2 | Not developed at the time of writing |
| **Modified scale** | | |
| Median (IQR) | 3.41 (2.99-3.64) | Not developed at the time of writing |
| Proportion rated poor (%, ≤2.5 score) | 7 | Not developed at the time of writing |
| **Subjective scale** | | |
| Median (IQR) | 2.50 (2.0-3.5) | Not developed at the time of writing |
| Proportion rated poor (%, ≤2.5 score) | 54 | Not developed at the time of writing |
| **Suitability Assessment of Material (%)** | | |
| Superior overall (70%-100%) | 44 | 15 |
| Adequate overall (40%-69%) | 53 | 39 |
| Not suitable (0%-39%) | 3 | 42 |
| **Readability, median (IQR)** | | |
| Flesch-Kincaid online tool | 8 (6-10) | 8 (7-10) |
| Flesch-Kincaid Word tool | 8 (6-9) | 8 (7-10) |
| Simple Measure of Gobbledygook | 8 (6-9) | 7 (7-8) |

# Discussion

## Principal Findings

This study evaluated 47 apps on infant feeding and physical activity and found that the information content within these apps was largely poor for coverage and depth of information presented. This study updated a systematic assessment of 46 apps available from 2013 to 2014, which similarly found poor quality information for parents [4]. Although the quality of apps, rated through the MARS, improved since the previous study, the credibility of advice did not improve; of the 59 evaluations, almost two-thirds reported *poor* coverage of information, and over half the app evaluations reported *incorrect or incomplete information* on the topics addressed.

Reviewers identified information contrary to the Australian guidelines on infant feeding [40] and physical activity [41] in 21 apps. Many of these apps were developed outside of Australia, that is, in America, the United Kingdom, and the European Union, where the official guidelines may be different from those in Australia. Although 1 app with incorrect information was developed in Australia and involved health care professionals, it was developed by a medical device company whose promotion of its breast pumps likely interfered with the provision of correct breastfeeding information.

The results using the MARS tool were largely positive. In this systematic assessment, the apps were rated as moderately engaging, functional, and visually appealing, and these apps clearly reported whether the information came from credible sources. One strength of the MARS tool is that it provides a multidimensional overview of app characteristics, all of which contribute to the user's experience of an app and the capacity to learn from it. In this case, most apps rated well on engagement, functionality, and aesthetics. However, in MARS, the quality and quantity of the information content constitute only 2 out of the 7 domains that contribute to the final score. Therefore, the information items were inadequate to reflect the importance of the content in this context. In other words, the MARS tool gave relatively little weight to the quality and accuracy of the information provided to consumers about infant nutrition and activity.

Therefore, although many apps were qualitatively evaluated as *good* to *excellent* on the objective and modified MARS score, the information contained within these apps was poor. This is reflected in the subjective scores for the apps, with a median of

2.50 and most apps falling between *inadequate* and *acceptable*. Reviewers reported that nearly half of all apps were *poor* or *inadequate* for recommending to others.

Brown et al [55,56] systematically reviewed the content of pregnancy apps against the Australian national recommendations on healthy eating for pregnancy. Similar to this study, they found that the apps were of moderate quality (mean objective MARS score was 3.05 [SD 0.66] and 3.52 [SD 0.58] for iOS and Android apps, respectively), with the functionality scale (mean 3.32 [SD 0.66]and 4.06 [SD 0.67], for iOS and Android apps, respectively) rated highest. These reviews also found incorrect and potentially harmful information in 3 apps about avoiding alcohol during pregnancy, fasting, restricting dairy products, and using dairy alternatives to meet calcium requirements.

Similarly, a review by Richardson et al [33] on apps for parents of infants in neonatal intensive care units highlighted the inconsistencies in quality ratings according to the method used. They identified a discrepancy between high app quality ratings using the MARS tool and those using the Trust It or Trash It tool [57] for evaluating clearly reported sources of health information.

Although most apps were rated suitable for their potential users on many dimensions, they scored less on readability and cultural appropriateness. Using the readability tools, only a handful of apps met the Australian government's requirement of year 4 reading level (3-5 apps, depending on what readability tool was used), although a majority met the less stringent South Australian government's requirement of year 8 level. However, the reviewers' SAM ratings suggested that most apps used a suitable vocabulary and appropriate style and ordering.

## Limitations

The app market is highly dynamic, and this study captured a cross-sectional snapshot of infant feeding and play apps available at one time point. Several apps in the previous study [4] were not available for download in 2018 and 2019. Even in this study, some apps had been removed from the marketplace by the time of writing, demonstrating the volatility of this information source. This changing availability makes the process of systematically reviewing apps challenging and potentially limited if key resources are not available during the selection phase.

Changes in the app search engines impacted the search process. The removal of the iOS App Store feature on iTunes in September 2017 [37] meant that the reviewers could only search using an iPhone instead of a desktop computer and prevented reliable double-checking of apps found for screening. As with other mHealth reviews conducted after September 2017 [56], double-checking of apps available in the App Store was conducted on Fnd, a web-based search tool [58], which reduces the flexibility of consumers to search for apps on desktop and synchronize the app download to their smartphone.

Search optimization, which refers to optimizing smartphone apps for visibility in search engines and browsing [59], also affected the types of apps found. To reflect the external validity of apps found by users searching through a smartphone app search engine, the authors did not include handsearching or searching for apps outside of the app search engine. Oversaturation of search engines with apps that are malware, spam, counterfeit, or contain copyrighted and farmed content [60-62] results in excessive number of apps; over 5000 apps were found during screening (Figure 1). This affects the apps found through Google Play, as search results are limited to 250 apps per search; therefore, if irrelevant apps are found in keyword searches and listed in the first 250 apps found, this is a nonoptimal search strategy, and relevant apps will be displaced from the search and not included in the analysis. This excessive number of apps caused relevant apps to be displaced by irrelevant apps; for example, a nongovernment organization–developed app on tummy time (*Red NoseSafe Sleeping*) was available on both search engines, but poor search optimization for keywords resulted in the app not being found through Google Play searches.

This study used keywords in English and evaluated apps available with English as the main language option. We acknowledge the multicultural background of the Australian population and the need for apps suitable for culturally and linguistically diverse parents [63], and a limitation of our study may exclude app users without English language proficiency (approximately 3.5% of Australians aged 15-49 years in 2016, data from Australian Bureau of Statistics [64]). Of the 102 relevant apps screened (Figure 1), only 3 were excluded for not being available in English. It is likely, however, that non-English–speaking parents may search for non-English sites. Zhao et al [32] used the 360 App Store and Chinese language keywords to conduct similar research, focusing on apps used by non-English–speaking parents. This shows the potential for this approach to include different app marketplaces and language search terms to explore apps available for different groups of parents.

## Interrater Reliability

The IRR results (Multimedia Appendix 4) showed considerable discrepancies on some instruments. There was low agreement on readability scores, which may reflect the different passages of text selected for evaluation by the pairs of reviewers. This is likely as reading grade level was calculated using objective tools, rather than individual reviewer assessment.

Although there was acceptable agreement on the modified MARS scores between the pairs of reviewers, there was far less agreement on the subjective MARS items. This difference in scoring may reflect the reviewer's perspectives (a dietitian researcher's evaluation compared with a child and family health nurse clinician's evaluation), affecting their likelihood of recommending an app and also their evaluation of its suitability for consumers using the SAM tool.

There was also mixed assessment on the coverage and depth of information on infant feeding, introduction to solids, and infant activity subtopics. Although reviewers evaluated content against the same government guidelines, they may have varied with regard to stringency in certain subtopics, such as deciding whether infants with appropriate signs of readiness can start solid foods between 4 and 6 months or if exclusive breastfeeding for 6 months was imperative [65]. Similarly, reviewers varied

on how much information constituted *complete* or *partially complete* advice, affecting IRR.

### Comparison With Prior Work

In replicating an earlier study of infant feeding and activity apps, we were able to compare the quality and content of apps available during 2018 to 2019 and those available 5 years earlier. The increased access to smartphones and the utilization of app-based health information among the Australian population over that time suggest that this study is highly relevant and timely. However, notwithstanding the increased suitability and functionality of these apps (measured with the SAM and MARS tools, respectively), the quality and accuracy of much of the information in apps have not greatly improved over time.

Since the 2015 study, this study identified growth in apps that are from reputable sources such as the government, universities, and health professionals: only 2 apps with university endorsements were found in the previous study [4] compared with 14 apps with university or government development or affiliation in this study. In addition, 3 apps were also used in research [52-54]. This finding indicates a positive transition of trustworthy sources that leverage the increased usage of technology and offer credible information to a wider population. This study also indicated that 9 apps were available in languages other than English, compared with none in the previous study. This demonstrates improved multilingual resources for a growing culturally diverse population.

This study was more comprehensive in that it looked at smartphone apps in Android and iOS, both paid and free to access. However, the previous study also looked at websites available on this topic, which might still be a widely used tool given the widespread use of *Dr Google* to search for pregnancy, birthing, and parenting information, which may not be accurate, credible, reliable, or safe [3,13,66-68].

Despite an increased proportion of apps published by reputable sources, this study also found relatively few good quality apps available. We reiterate the earlier recommendation [5] to establish a certified endorsement for apps similar to the Health on the Net Foundation Code of Conduct used on the website. This code of conduct is used to standardize the reliability of medical and health information available on the World Wide Web [69] and encourages website developers to maintain the quality standards of the organization. The 2015 study found that websites that subscribed to this code of conduct certificate had higher quality scores [4].

Only 2 apps (*WebMD Baby* and *Pregnancy and Baby Tracker* [formerly *What to Expect*]) were included across both studies, highlighting that the app marketplace changes continuously. A potential reason for the short *shelf life* of some apps could be the high maintenance cost required to keep the apps updated with the evolving smartphone operating systems. Intervention studies of health apps have reported technical issues with the implementation of their app [70], such as updates in the operating system or app impeding participant access [71-74]. This indicates that app functioning requires ongoing maintenance. This might be a challenge for apps that are developed with limited funding from universities or governments compared with commercial companies that often have higher budgets. Further research is required to explore factors that impact the shelf life of health-related apps and to develop suggestions on sustainable ways to leverage technology to share health-related information.

The use of the MARS tool, specifically developed for app evaluation, is a strength of this study. However, the MARS tool had not been published at the time of the 2015 study [4]. The original study adapted the Quality Component Scoring System to assess apps; however, this tool was originally developed to evaluate the quality of medical websites [75], and not all items were appropriate for apps.

Although the national infant feeding guidelines in Australia have remained consistent since the previous study [40], the uptake of this information has not improved. Research indicates that gaps exist in the current infant feeding practices. Begley et al's [76] research with mothers, using focus groups, in Western Australia found that less than half of the participants had heard of the Australian Infant Feeding Guidelines or were aware that the recommended age for the introduction of solid foods was around 6 months; many participants believed that the guidelines were based on opinion rather than scientific research. A survey of mother-infant dyads in Western Australia and South Australia during 2010 to 2011 found that the feeding behaviors of participants fell short of Australian feeding guidelines, where although 93% of mothers initiated breastfeeding, only 42% of infants were breastfed to 6 months, and 97% of infants received solid food by 6 months [77].

### Implications for Practice

It is well established that early childhood experiences have a significant impact on optimal child development [40]. Child and family health nurses in the community play an integral role in monitoring children's growth and development and providing guidance and support to parents. Child and family health nurses and lactation consultants work in an environment that has limited staffing and financial resources.

Many parents live in isolation from extended families and turn to social networking sites for advice from peers for health-related information. Conflicting advice and lack of continuity of care from health professionals often adds to their confusion about how they should care for their children [78,79]. Increasingly, parents require direction and guidance to seek evidence-based educational resources. Parents seek information that is easily accessible and affordable. Information provided during consultation in the child and family health nurse clinics often dispels the parents' concerns, but parents may be unable to absorb all the information at one time and often require educational resources that they can refer back to once they have gone home. Appropriate internet websites and smartphone apps are key to meeting this need.

Parents, particularly first-time parents, are bombarded with information from a variety of sources, especially from websites and smartphone apps. Although information evaluation tools supported by librarians, academics, and the government [57,80,81] for critically assessing health information quality can support the health and digital literacy of parents, these tools

may not be widely known to the layperson. Similarly, clinicians who regularly support new parents often receive requests for advice about apps and cannot confidently provide a recommendation if they are unaware of app evaluation tools [82] or have insufficient time to evaluate apps. One way to overcome this challenge is to establish a *trusted app* or similar logo (logo similar to that of the Health on the Net Foundation Code of Conduct's logo) or a repository of approved apps, such as the United Kingdom's National Health Service Apps Library [83], which can be applied to evidence-based apps that do not promote particular products.

## Conclusions

Improved functionality, suitability, and user engagement of smartphone apps in recent years are welcome developments for parents seeking guidance on infant feeding and activity. However, the high-quality content of apps is critical to good health outcomes. Assessment of available apps revealed that some provided very useful information, but there was wide disparity in reliability and consistency with evidence-based knowledge. Many apps provided additional information outside their focal topic or area of expertise; others offered information largely designed to promote their products. In some instances, this information was limited and did not provide comprehensive advice consistent with evidence-based guidelines.

## Authors' Contributions

EDW and ST designed the study and the main conceptual ideas. AT, CL, DS, HC, JJ, and EDW undertook the app searches on Google Play search engine or iOS App Store and evaluated the apps. HC undertook statistical analysis. HC, CR, EDW, and ST wrote the paper with input from all authors.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Google Trends for app search key terms.
[DOCX File , 57 KB - mhealth_v8i5e17300_app1.docx ]

Multimedia Appendix 2
App search and evaluation protocol.
[DOCX File , 4376 KB - mhealth_v8i5e17300_app2.docx ]

Multimedia Appendix 3
App evaluation tools used.
[PDF File (Adobe PDF File), 407 KB - mhealth_v8i5e17300_app3.pdf ]

Multimedia Appendix 4
Supplementary tables and figures.
[DOCX File , 372 KB - mhealth_v8i5e17300_app4.docx ]

## References

1. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Nov;294(5):937-944. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]
2. Kennedy RA, Mullaney L, Reynolds CM, Cawley S, McCartney DM, Turner MJ. Preferences of women for web-based nutritional information in pregnancy. Public Health 2017 Feb;143:71-77. [doi: 10.1016/j.puhe.2016.10.028] [Medline: 28159029]
3. Walker LO, Mackert MS, Ahn J, Vaughan MW, Sterling BS, Guy S, et al. e-Health and new moms: contextual factors associated with sources of health information. Public Health Nurs 2017 Nov;34(6):561-568. [doi: 10.1111/phn.12347] [Medline: 28762533]
4. Taki S, Campbell KJ, Russell CG, Elliott R, Laws R, Denney-Wilson E. Infant feeding websites and apps: a systematic assessment of quality and content. Interact J Med Res 2015 Sep 29;4(3):e18 [FREE Full text] [doi: 10.2196/ijmr.4323] [Medline: 26420339]
5. Silver L. Pew Research Center. Washington, DC, USA: Pew Research Center; 2019 Feb 5. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally URL: http://www.pewglobal.org/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-03-18] [WebCite Cache ID 76xMvF7eM]

6. Australian Bureau of Statistics. Canberra, Australian Capital Territory, Australia: Australian Government; 2018 Mar 28. 8146.0 - Household Use of Information Technology, Australia, 2016-17 URL: https://www.abs.gov.au/ausstats/abs@.nsf/mf/8146.0 [accessed 2019-03-20] [WebCite Cache ID 770Q9SJt3]
7. Australian Communications and Media Authority. Canberra, Australian Capital Territory, Australia: Australian Government; 2017 Nov. Communications Report 2016-17 URL: https://www.acma.gov.au/publications/2017-09/report/communications-report-2016-17 [accessed 2019-03-18] [WebCite Cache ID 76xNRFCKr]
8. Corbett P, Huggins K. Deloitte Australia.: Deloitte Touche Tohmatsu; 2018. Mobile Consumer Survey 2019 - The Australian Cut URL: https://www2.deloitte.com/au/mobileconsumer [accessed 2019-03-18] [WebCite Cache ID 76xOHoJqq]
9. Schuman AJ. What's new in baby tech. Contemp Pediatr 2016;37(3):37-40 [FREE Full text]
10. Wellde PT, Miller LA. There's an app for that!: new directions using social media in patient education and support. J Perinat Neonatal Nurs 2016;30(3):198-203. [doi: 10.1097/JPN.0000000000000177] [Medline: 27465449]
11. Aungst TD, Clauson KA, Misra S, Lewis TL, Husain I. How to identify, assess and utilise mobile medical applications in clinical practice. Int J Clin Pract 2014 Feb;68(2):155-162. [doi: 10.1111/ijcp.12375] [Medline: 24460614]
12. Chaudhry BM. Expecting great expectations when expecting. Mhealth 2018;4:2 [FREE Full text] [doi: 10.21037/mhealth.2017.12.01] [Medline: 29445731]
13. Pehora C, Gajaria N, Stoute M, Fracassa S, Serebale-O'Sullivan R, Matava CT. Are parents getting it right? A survey of parents' internet use for children's health care information. Interact J Med Res 2015 Jun 22;4(2):e12 [FREE Full text] [doi: 10.2196/ijmr.3790] [Medline: 26099207]
14. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [doi: 10.1071/PY13039] [Medline: 23899373]
15. Helle C, Hillesund ER, Wills AK, Øverby NC. Evaluation of an ehealth intervention aiming to promote healthy food habits from infancy-the Norwegian randomized controlled trial early food for future health. Int J Behav Nutr Phys Act 2019 Jan 3;16(1):1 [FREE Full text] [doi: 10.1186/s12966-018-0763-4] [Medline: 30606197]
16. Australian Breastfeeding Association, Reach Health Promotion Innovations, Curtin University, Healthway. Feed Safe. 2016. URL: http://www.feedsafe.net/ [accessed 2019-03-18] [WebCite Cache ID 76xQmY5ni]
17. NHS. National Health Service UK. Baby Buddy URL: https://www.nhs.uk/apps-library/baby-buddy/ [accessed 2019-03-18] [WebCite Cache ID 76xR2haTz]
18. Family Health Service.: The Government of Hong Kong Special Administrative Region; 2016. 'Info for Nursing Mum' Mobile App URL: https://www.fhs.gov.hk/tc_chi/archive/files/bf_app.pdf [accessed 2019-03-18] [WebCite Cache ID 76xRTowrK]
19. World Vision Philippines. Quezon City, Philippines: World Vision; 2018. Mother-Baby Friendly Philippines to Turn Over its Reporting Platform to DOH URL: https://www.worldvision.org.ph/stories/mother-baby-friendly-philippines-to-turn-over-its-reporting-platform-to-doh/ [accessed 2019-03-18] [WebCite Cache ID 76xRi44zz]
20. Midland Maternity Action Group, HealthShare Ltd, Midland District Health Boards. BreastFedNZ. 2018. URL: http://www.breastfednz.co.nz/ [accessed 2019-03-18] [WebCite Cache ID 76xRynAHS]
21. Schindler-Ruwisch JM, Roess A, Robert RC, Napolitano MA, Chiang S. Social support for breastfeeding in the era of mhealth: a content analysis. J Hum Lact 2018 Aug;34(3):543-555. [doi: 10.1177/0890334418773302] [Medline: 29787686]
22. Jimenez-Zabala A. Clinical Trials.: US National Library of Medicine; 2016. Mother's Milk Messaging: Evaluation of a Bilingual APP to Support Initiation and Exclusive Breastfeeding in New Mothers (MMM) URL: https://clinicaltrials.gov/ct2/show/NCT02958475 [accessed 2019-03-18] [WebCite Cache ID 76xOkkwZ9]
23. White BK, Martin A, White JA, Burns SK, Maycock BR, Giglia RC, et al. Theory-based design and development of a socially connected, gamified mobile app for men about breastfeeding (Milk Man). JMIR Mhealth Uhealth 2016 Jun 27;4(2):e81 [FREE Full text] [doi: 10.2196/mhealth.5652] [Medline: 27349756]
24. Lewkowitz AK, Anderson M. Clinical Trials.: US National Library of Medicine; 2017. Impact of a Smartphone Application on Postpartum Weight Loss and Breastfeeding Rates Among Low-income, Urban Women URL: https://clinicaltrials.gov/ct2/show/NCT03167073 [accessed 2019-03-18] [WebCite Cache ID 76xPH9QJ5]
25. Tang K, Gerling K, Chen W, Geurts L. Information and communication systems to tackle barriers to breastfeeding: systematic search and review. J Med Internet Res 2019 Sep 27;21(9):e13947 [FREE Full text] [doi: 10.2196/13947] [Medline: 31573903]
26. Resourcing Parents.: Families NSW; 2019. Love Talk Sing Read Play App URL: http://ltsrp.resourcingparents.nsw.gov.au/home/resources [accessed 2019-04-10] [WebCite Cache ID 77XZlExuL]
27. Deadly Tots.: Families NSW; 2019. Deadly Tots App URL: http://deadlytots.com.au/Page/deadlytotsapp [accessed 2019-04-10] [WebCite Cache ID 77Xa1kxFS]
28. NSW Health.: NSW Government; 2019. New App Helps On-Time Vaccination for Kids URL: https://www.health.nsw.gov.au/news/Pages/20181210_00.aspx [accessed 2019-04-11] [WebCite Cache ID 77Xa9llZQ]
29. Red Nose Australia. healthdirect. 2019. Red Nose Safe Sleeping App URL: https://www.healthdirect.gov.au/red-nose-safe-sleeping-app [accessed 2019-04-10] [WebCite Cache ID 77XaPDY2z]

30. Wen LM, Baur LA, Simpson JM, Rissel C, Wardle K, Flood VM. Effectiveness of home based early intervention on children's BMI at age 2: randomised controlled trial. Br Med J 2012 Jun 26;344:e3732 [FREE Full text] [doi: 10.1136/bmj.e3732] [Medline: 22735103]
31. NSW Health.: NSW Government; 2018. $1.2 Million Bump for Healthy Mums and Bubs URL: https://www.health.nsw.gov.au/news/Pages/20180723_01.aspx [accessed 2019-04-10] [WebCite Cache ID 77XYDxp3H]
32. Zhao J, Freeman B, Li M. How do infant feeding apps in China measure up? A content quality assessment. JMIR Mhealth Uhealth 2017 Dec 6;5(12):e186 [FREE Full text] [doi: 10.2196/mhealth.8764] [Medline: 29212627]
33. Richardson B, Dol J, Rutledge K, Monaghan J, Orovec A, Howie K, et al. Evaluation of mobile apps targeted to parents of infants in the neonatal intensive care unit: systematic app review. JMIR Mhealth Uhealth 2019 Apr 15;7(4):e11620 [FREE Full text] [doi: 10.2196/11620] [Medline: 30985282]
34. Australian Communications and Media Authority. Canberra, Australian Capital Territory, Australia: Australian Government; 2013. Communications Report 2012-13 URL: https://www.acma.gov.au/sites/default/files/2019-08/Communications-report-2012-13.pdf [accessed 2019-03-20] [WebCite Cache ID 770OQRJ6k]
35. Research 2 Guidance. Berlin, Germany: Research2Guidance; 2017. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2019-03-20] [WebCite Cache ID 770RTnz1k]
36. Google Play. California, USA: Google LLC; 2019. Android Apps on Google Play URL: https://play.google.com/store/apps [accessed 2019-04-10]
37. Smith J. ZDNet.: CBS Interactive; 2017 Sep 13. Apple's iTunes Removes iOS App Store From Desktop Version URL: https://www.zdnet.com/article/apple-removes-ios-app-store-from-desktop-versions-of-itunes/ [accessed 2019-01-23] [WebCite Cache ID 75dI5ofhb]
38. Carr A. Research Guides - Austin Community College. Austin, Texas, USA: Austin Community College; 2019. Content Farms: What is a Content Farm? URL: https://researchguides.austincc.edu/contentfarms [accessed 2019-09-19]
39. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
40. National Health and Medical Research Council. Infant Feeding Guidelines: Information for Health Workers. Canberra, Australian Capital Territory, Australia: Australian Government; 2012. URL: https://www.eatforhealth.gov.au/sites/default/files/files/the_guidelines/n56_infant_feeding_guidelines.pdf [accessed 2019-01-21]
41. Department of Health. Canberra, Australian Capital Territory, Australia: Australian Government; 2010. Australian 24-Hour Movement Guidelines for the Early Years (Birth to 5 Years) URL: http://www.health.gov.au/internet/main/publishing.nsf/Content/F01F92328EDADA5BCA257BF0001E720D/$File/Birthto5years_24hrGuidelines_Brochure.pdf [accessed 2019-01-21] [WebCite Cache ID 75dJf5jqH]
42. Health on the Net Foundation. Per Carlbring: Nyheter. Health-Related Web Site Evaluation Form URL: https://www.carlbring.se/form/itform_eng.pdf [accessed 2020-01-21] [WebCite Cache ID 75dJp4qhc]
43. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
44. Doak C, Doak L, Root J. Aspirius Library. 2008. Suitability Assessment of Materials For Evaluation of Health-Related Information for Adults URL: http://aspiruslibrary.org/literacy/sam.pdf [accessed 2019-01-23] [WebCite Cache ID 75dJtwmUM]
45. Australian Government. Digital Guides. Canberra, Australian Capital Territory, Australia: Commonwealth of Australia; 2019. Writing Style URL: https://guides.service.gov.au/content-guide/writing-style/ [accessed 2019-01-23] [WebCite Cache ID 75dJyG9FZ]
46. Kincaid PJ, Fishbourne Jr R, Rogers R, Chissom BS. ucf stars - University of Central Florida. Milington, Tennessee, USA: Institute for Simulation and Training; 1975. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel URL: https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1055&context=istlibrary [accessed 2019-01-21] [WebCite Cache ID 75dKAn3hb]
47. McLaughlin HG. SMOG Grading - a new readability formula. J Read 1969;12(8):639-646 [FREE Full text]
48. My Byline Media. Readability Formulas. URL: http://www.readabilityformulas.com/ [accessed 2019-01-23] [WebCite Cache ID 75dKOmJzE]
49. Microsoft Word Help Center. Microsoft Support. Washington, USA: Microsoft Corporation; 2019. Get Your Document's Readability and Level Statistics URL: https://support.office.com/en-us/article/test-your-document-s-readability-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2 [accessed 2019-01-23] [WebCite Cache ID 75dKX8uA8]
50. SA Health. Adelaide, South Australia, Australia: Government of South Australia; 2013. Policy Guideline: Guide for Engaging with Consumers and the Community URL: https://tinyurl.com/y92z2o54 [accessed 2019-01-21] [WebCite Cache ID 777ir6hwR]

51. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Second Edition. Thousand Oaks, CA: Sage Publications; 2014.
52. Manlongat D. University of Windsor Electronic Theses and Dissertations. Ontario, Canada: University of Windsor; 2017. The Effects of Introducing Prenatal Breastfeeding Education in the Obstetricians' Waiting Rooms URL: https://scholar.uwindsor.ca/etd/7276/ [accessed 2020-04-02]
53. Wheaton N, Lenehan J, Amir LH. Evaluation of a breastfeeding app in rural Australia: prospective cohort study. J Hum Lact 2018 Nov;34(4):711-720. [doi: 10.1177/0890334418794181] [Medline: 30235043]
54. Taki S, Russell CG, Wen LM, Laws RA, Campbell K, Xu H, et al. Consumer engagement in mobile application (app) interventions focused on supporting infant feeding practices for early prevention of childhood obesity. Front Public Health 2019;7:60 [FREE Full text] [doi: 10.3389/fpubh.2019.00060] [Medline: 30984732]
55. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy apps freely available in the Google Play Store. Health Promot J Austr 2019 Jun 21 epub ahead of print. [doi: 10.1002/hpja.270] [Medline: 31225924]
56. Brown HM, Bucher T, Collins CE, Rollo ME. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. Matern Child Nutr 2019 Jul;15(3):e12768. [doi: 10.1111/mcn.12768] [Medline: 30569549]
57. Genetic Alliance. Trust It or Trash It?. Washington DC, USA: Access To Credible Genetics Resource Network; 2009. URL: http://www.trustortrash.org/TrustorTrash.pdf [accessed 2019-09-06]
58. Mack J, Mack R. Fnd. 2019. URL: https://fnd.io/ [accessed 2019-09-16]
59. Samuel S. Search engine optimisation to improve your visibility online. In Pract 2013 Jun 17;35(6):346-349. [doi: 10.1136/inp.f2703]
60. Ahn A. Android Developers Blog. California, USA: Google LLC; 2018 Jan 30. How We Fought Bad Apps and Malicious Developers in 2017 URL: https://android-developers.googleblog.com/2018/01/how-we-fought-bad-apps-and-malicious.html [accessed 2019-09-05]
61. Seneviratne S, Seneviratne A, Kaafar MA, Mahanti A, Mohapatra P. Spam mobile apps: characteristics, detection, and in the wild analysis. ACM Trans Web 2017 Apr 10;11(1):1-29. [doi: 10.1145/3007901]
62. Wang H, Li H, Li L, Guo Y, Xu G. Why Are Android Apps Removed From Google Play?: A Large-Scale Empirical Study. In: Proceedings of the 15th International Conference on Mining Software Repositories. 2018 Presented at: MSR'18; May 28-29, 2018; Gothenburg, Sweden p. 231-242. [doi: 10.1145/3196398.3196412]
63. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. JMIR Mhealth Uhealth 2018 Nov 16;6(11):e189 [FREE Full text] [doi: 10.2196/mhealth.9119] [Medline: 30446483]
64. Australian Bureau of Statistics. Canberra, Australian Capital Territory, Australia: Australian Government; 2016. Proficiency in Spoken English/Language by Age by Sex (SA2+) URL: http://stat.data.abs.gov.au/Index.aspx?DataSetCode=ABS_C16_T08_SA [accessed 2020-03-03]
65. Hetherington MM, Cecil JE, Jackson DM, Schwartz C. Feeding infants and young children. From guidelines to practice. Appetite 2011 Dec;57(3):791-795. [doi: 10.1016/j.appet.2011.07.005] [Medline: 21784114]
66. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016 Jul 19;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
67. Fleming SE, Vandermause R, Shaw M. First-time mothers preparing for birthing in an electronic world: internet and mobile phone technology. J Reprod Infant Psychol 2014 Mar 19;32(3):240-253. [doi: 10.1080/02646838.2014.886104]
68. Lynch C, Nikolova G. The internet: a reliable source for pregnancy and birth planning? A qualitative study. MIDIRS Midwifery Dig 2015;25(2):193-199 [FREE Full text]
69. Health On the Net. Geneva, Switzerland: Health on the Net Foundation; 2019. The Commitment to Reliable Health and Medical Information on the Internet URL: https://www.hon.ch/HONcode/Patients/Visitor/visitor.html [accessed 2019-11-25]
70. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
71. White B, Giglia RC, White JA, Dhaliwal S, Burns SK, Scott JA. Gamifying breastfeeding for fathers: process evaluation of the Milk Man mobile app. JMIR Pediatr Parent 2019 Jun 20;2(1):e12157 [FREE Full text] [doi: 10.2196/12157] [Medline: 31518324]
72. Litterbach E, Russell CG, Taki S, Denney-Wilson E, Campbell KJ, Laws RA. Factors influencing engagement and behavioral determinants of infant feeding in an mHealth program: qualitative evaluation of the Growing healthy program. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e196 [FREE Full text] [doi: 10.2196/mhealth.8515] [Medline: 29254908]
73. Taki S, Russell CG, Lymer S, Laws R, Campbell K, Appleton J, et al. A mixed methods study to explore the effects of program design elements and participant characteristics on parents' engagement with an mHealth program to promote healthy infant feeding: the Growing healthy program. Front Endocrinol (Lausanne) 2019;10:397 [FREE Full text] [doi: 10.3389/fendo.2019.00397] [Medline: 31293515]

74. Venter W, Coleman J, Chan VL, Shubber Z, Phatsoane M, Gorgens M, et al. Improving linkage to HIV care through mobile phone apps: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 17;6(7):e155 [FREE Full text] [doi: 10.2196/mhealth.8376] [Medline: 30021706]
75. Martins EN, Morse LS. Evaluation of internet websites about retinopathy of prematurity patient education. Br J Ophthalmol 2005 May;89(5):565-568 [FREE Full text] [doi: 10.1136/bjo.2004.055111] [Medline: 15834086]
76. Begley A, Ringrose K, Giglia R, Scott J. Mothers' understanding of infant feeding guidelines and their associated practices: a qualitative analysis. Int J Environ Res Public Health 2019 Mar 29;16(7):1141 [FREE Full text] [doi: 10.3390/ijerph16071141] [Medline: 30934967]
77. Magarey A, Kavian F, Scott JA, Markow K, Daniels L. Feeding mode of Australian infants in the first 12 months of life. J Hum Lact 2016 Nov;32(4):NP95-N104. [doi: 10.1177/0890334415605835] [Medline: 26416440]
78. Hesson A, Fowler C, Rossiter C, Schmied V. 'Lost and confused': parent representative groups' perspectives on child and family health services in Australia. Aust J Prim Health 2017 Dec;23(6):560-566. [doi: 10.1071/PY17072] [Medline: 29183564]
79. Rossiter C, Fowler C, Hesson A, Kruske S, Homer CS, Kemp L, et al. Australian parents' experiences with universal child and family health services. Collegian 2019 Jun;26(3):321-328. [doi: 10.1016/j.colegn.2018.09.002]
80. Meriam Library. Chico, California, USA: California State University, Chico; 2004. Evaluating Information – Applying the CRAAP Test URL: https://library.csuchico.edu/sites/default/files/craap-test.pdf [accessed 2020-03-02]
81. US National Library of Medicine, US Department of Health and Human Services, National Institutes of Health. Medline Plus. Bethesda, Maryland, USA: US National Library of Medicine; 2017. Evaluating Health Information URL: https://medlineplus.gov/evaluatinghealthinformation.html [accessed 2020-03-02]
82. Skiba D. Evaluation tools to appraise social media and mobile applications. Inform 2017 Sep 15;4(3):32. [doi: 10.3390/informatics4030032]
83. NHS. National Health Service UK. 2019. NHS Apps Library URL: https://www.nhs.uk/apps-library/ [accessed 2019-09-06]


## Abbreviations

**F-K:** Flesch-Kincaid
**iOS:** iPhone Operating System
**IRR:** interrater reliability
**MARS:** Mobile App Rating Scale
**mHealth:** mobile health
**SAM:** Suitability Assessment of Materials
**SMOG:** Simple Measure of Gobbledygook


*Edited by G Eysenbach; submitted 05.12.19; peer-reviewed by J Ropero, L Levin; comments to author 28.02.20; revised version received 10.03.20; accepted 12.03.20; published 27.05.20.*

*Please cite as:*
*Cheng H, Tutt A, Llewellyn C, Size D, Jones J, Taki S, Rossiter C, Denney-Wilson E*
*Content and Quality of Infant Feeding Smartphone Apps: Five-Year Update on a Systematic Search and Evaluation*
*JMIR Mhealth Uhealth 2020;8(5):e17300*
*URL: http://mhealth.jmir.org/2020/5/e17300/*
*doi:10.2196/17300*
*PMID:32459187*

Original Paper

# Mobile Apps for Mental Health Issues: Meta-Review of Meta-Analyses

Tania Lecomte[1,2], PhD; Stéphane Potvin[2,3], PhD; Marc Corbière[2,4], PhD; Stéphane Guay[2,5], PhD; Crystal Samson[1], BSc; Briana Cloutier[1], BSc; Audrey Francoeur[1], MSc; Antoine Pennou[1], BSc; Yasser Khazaal[6], MD

[1]Department of Psychology, University of Montreal, Montreal, QC, Canada

[2]Centre de recherche, l'Institut Universitaire en Santé Mentale de Montréal, Montreal, QC, Canada

[3]Department of Psychiatry, University of Montreal, Montreal, QC, Canada

[4]Department of Education, Career Counselling, University du Quebec a Montreal, Montreal, QC, Canada

[5]Department of Criminology, University of Montreal, Montreal, QC, Canada

[6]Department of Psychiatry, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland

**Corresponding Author:**
Tania Lecomte, PhD
Department of Psychology
University of Montreal
90 rue Vincent d'Indy
Bur C-358, CP 6128, Succ Centre-Ville
Montreal, QC, H3C 3J7
Canada
Phone: 1 343 6274
Email: tania.lecomte@umontreal.ca

## Abstract

**Background:** Mental health apps have great potential to help people needing support to cope with distress or specific symptoms. In fact, there is an exponential increase in the number of mental health apps available on the internet, with less than 5% being actually studied.

**Objective:** This study aimed to assess the quality of the available evidence regarding the use of mental health apps and to summarize the results obtained so far.

**Methods:** Systematic reviews and meta-analyses were searched, specifically for mobile apps on mental health issues or symptoms, and rated using the Grading of Recommendations Assessment, Development and Evaluation system.

**Results:** A total of 7 meta-analyses were carefully reviewed and rated. Although some meta-analyses looked at any mental health issue and analyzed the data together, these studies were of poorer quality and did not offer strong empirical support for the apps. Studies focusing specifically on anxiety symptoms or depressive symptoms were of moderate to high quality and generally had small to medium effect sizes. Similarly, the effects of apps on stress and quality of life tended to offer small to medium effects and were of moderate to high quality. Studies looking at stand-alone apps had smaller effect sizes but better empirical quality than studies looking at apps with guidance. The studies that included follow-ups mostly found a sustained impact of the app at an 11-week follow-up.

**Conclusions:** This meta-review revealed that apps for anxiety and depression hold great promise with clear clinical advantages, either as stand-alone self-management or as adjunctive treatments. More meta-analyses and more quality studies are needed to recommend apps for other mental health issues or for specific populations.

***(JMIR Mhealth Uhealth 2020;8(5):e17458)*** doi:10.2196/17458

**KEYWORDS**

apps; mental health; depression; anxiety; review; meta

# Introduction

## Mobile Health and Apps

Recent years have seen an exponential development of mobile technologies aimed at improving various mental health problems. Such technologies are considered part of a new field of medicine called mobile health (*mHealth*). This term refers to health (including mental health) supported by mobile technologies [1]. The mHealth field is booming, with a plethora of health-related apps, websites, and text messaging–support interventions being developed by the industry and being adopted by the public [2]. However, only a small proportion of these technologies have undergone any form of empirical assessment [3].

This lack of app validation is a concern, even more so when studies suggest that mental health– and addiction-related apps currently available to the public, with few exceptions, offer insufficient content quality [4-8]. Fortunately, recent years have seen an increase in the gathering of empirical data related to smartphone app–related interventions [9].

## What Are Mobile Mental Health Interventions?

According to the World Health Organization's definition of mHealth, mobile mental health interventions could be considered as mental health services (medical and public health practices) "supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants (PDAs), and other wireless devices" [10]. They include smartphone apps, voice, video or text messaging interventions, real-time tracking, and Web-based interventions, to name a few. In this review, we were specifically interested in mental health apps.

Smartphone apps, because of their worldwide mobility, connectivity, 24-hour availability, and their ubiquitous characteristics, are strong vectors for mHealth interventions. Furthermore, they can convey a large range of technologies and functionalities, such as virtual reality, augmented reality (inserting computer elements into the real field), telemedicine, robotics, games, interfaces connected to sensors, social networks, real-time interactivity, geolocation, and more [11].

App technology has shown the greatest reach in the past few years. According to the United Nations, more than 90% of the population in developed countries use such apps daily [12]. Once available for the target audience, because of the wide dissemination of smartphone devices, such tools can attract many downloads from all over the world. This potential is illustrated by tens of thousands of downloads of such apps [13,14].

The ubiquitous, handy mobile format and 24-hour availability of smartphones offer an important advantage for using apps to target mental health problems. One may hypothesize that the treatment of mental disorders could be improved by effective support in the right place at the right time. As learning is context dependent [15], apps can support the process of empowerment and recovery of people with various mental health problems by allowing people to access tools or support when needed [16,17].

Mental health apps can also offer the opportunity to assess, with Ecological Momentary Assessment, or intervene, via Ecological Momentary Interventions (EMIs), individuals in their natural environment, thereby enabling a better understanding of the factors triggering problems and addressing the problems when and where they arise [18-20]. These methods overcome memory biases by asking questions pertaining to the current moment or the current day and can also help determine if phenomena are stable or change from day to day [21]. Furthermore, such ecologically valid data may help to guide treatments or improve assessments in naturalistic settings [22].

## Current Knowledge on Mobile Apps for Mental Health Problems

There is currently somewhere between 10,000 and 20,000 *mental health* apps [23,24], but it is estimated that only about 3% to 4% are actually evidence based. Most of these studies have been conducted in the last few years and assessed either the feasibility and acceptability of mental health apps and, in some cases, their efficacy for a broad spectrum of mental disorders, including depressive disorders, posttraumatic stress disorders, schizophrenia, bipolar disorders, or addictions [9,14,25-38].

Given the heterogeneity and speed of publication of app-related studies, aggregated results are needed to determine the overall (vs specific) efficacy of mobile apps for mental health. Multiple meta-analyses on apps focusing on a single or multiple mental health problems have been conducted [39,40], with very different results at times. This could be explained by the selection criteria for the meta-analyses, with some only focusing on stand-alone apps, others only looking at adjunctive apps (apps offered on top of another treatment) or apps offered with guidance (a person available for questions or to prompt its use), others considering both models together, and others still including everything and evaluating the models separately in different subanalyses. In fact, some authors suggest that only adjunctive apps or apps with guidance should be recommended at this point for mental health issues [41]. Given the speed of uptake of many of these apps, it is important to determine, based on the quality of the evidence available and the effect sizes, if we should recommend such apps for mental health problems such as depression or anxiety. The purpose of this meta-review was to summarize these results and determine the empirical quality of the evidence reported using the grading of recommendations, assessment, development, and evaluation (GRADE) system [42]. This system permits the quality of evidence produced by meta-analyses to be evaluated, according to specific factors: the sample size, the stable findings across studies, the appropriate control for known confounding factors, no evidence of study bias, follow-up (if any), and results being closely linked to the outcomes targeted here (see Tables 1 and 2). The GRADE system has been successfully applied to meta-analyses of pre-post designs, randomized controlled trials (RCTs), correlational studies, experimental studies, and longitudinal studies [42].

**Table 1.** GRADE review of included meta-analyses.

| Authors | Technology used | Intervention type and subtype or outcome type | Effect size | Precise (less than 0.25 = precise) | Consistent | Direct |
|---|---|---|---|---|---|---|
| | | | d, g, OR, RR | 95% CI, Yes or No | Q or $I^2$, Yes or No | Yes or No |
| **Lindhiem et al., 2015** | mHealth[a] for psychotherapy or behavioural interventions (across all problems and issues) | Specifically for Apps (excluding PDAs), all mental health problems together | d= 0.57 | 95% CI: 0.28-0.85 - No | Q= 125.15, *P*>.001, (overall, not specific for apps) | No (multiple outcomes combined) |
| **Versluis et al., 2016** | Ecological momentary interventions (EMI) for anxiety, depression, stress and positive mental health | Ecological Momentary Interventions (EMIs); outcome type : global mental health, outliers removed | g=0.57 | 95% CI: 0.45-0.70, Yes | $I^2$: 65.08, No | No (multiple outcomes) |
| | N/A[b] | EMIs; outcome type : global mental health compared to control conditions | g=0.65 | 95% CI : 0.48-0.82, No | $I^2$: 58.5, No | No |
| | N/A | EMIs; outcome: anxiety | g=0.47 | 95% CI: 0.32-0.63 No | $I^2$=50.48; No | Yes |
| | N/A | EMIs; outcome : depression | g=0.48 | 95% CI: 0.34-0.61 No | $I^2$=65.58; No | Yes |
| | N/A | EMIs; outcome : perceived stress | g=0.40 | 95% CI : 0.23-0.57 No | $I^2$= 12.79; Yes | No |
| | N/A | EMIs; outcome : quality of life | g= 0.38 | 95% CI : 0.19-0.56 No | $I^2$=0; Yes | No |
| | N/A | with guidance | g=0.73 | 95% CI : 0.57-0.88 Yes | $I^2$=37.1% Q=20.67; Yes | Yes |
| | N/A | stand alone | g= 0.45 | 95% CI : 0.22-0.69 No | $I^2$=77.7% Q=36.8, *P*=.05; No | Yes |
| **Stratton et al., 2017** | eHealth and mHealth (app or web-based) for mental health at work | eHealth in the work place; outcome : global (No specific analyses just for apps) | g=0.24 | 95% CI: 0.13-0.35 (No) | $I^2$: 67.6%, Q = 9.82 (df2), *P*<=.01, No | No (eHealth[c] Smartphone; multiple outcomes) |
| **Firth et al., 2017a** | Psychological interventions for anxiety via smartphones | Smartphone interventions, all studies, anxiety symptoms | g=0.33 | 95% CI: 0.17-0.48 (No) | Q=15.9, $I^2$=49.6%), No | Yes (anx) |
| | N/A | compared to active controls | g=0.19 | 0.06-0.31, Yes | Yes | Yes |
| | N/A | compared to waitlist | g=0.45 | 0.3-0.6, No | Yes | Yes |
| **Firth et al., 2017b** | Psychological interventions for depression via smartphones | Smartphone interventions for depression, global | g = 0.38 | 95% CI: 0.242 - 0.524, Yes | Q = 80.8 *P*=.01 $I^2$=74%, No | Yes |
| | N/A | compared to active controls | g = 0.22 | 95% CI: 0.098 - 0.334, Yes | Q = 20.8 *P*=.03 $I^2$=47.2, No | Yes |

| Authors | Technology used | Intervention type and subtype or outcome type | Effect size | Precise (less than 0.25 = precise) | Consistent | Direct |
|---|---|---|---|---|---|---|
| | | | d, g, OR, RR | 95% CI, Yes or No | Q or $I^2$, Yes or No | Yes or No |
| | | compared to inactive controls | g = 0.56 | 95% CI: 0.379 - 0.736, No | Q =34.9 *P*=.01 $I^2$=65.6, No | Yes |
| **Witt et al., 2017** | eHealth and mHealth (app or web-based) for self-management of suicidal ideation and self-harm – (stand alone only) | suicidal ideation RCT[d] | d=-0.26 | 95% CI: -0.44- -0.08, No | $I^2$: 0%, Yes | Yes (suicidal ideation) |
| | N/A | suicidation ideation (No controls) | d=-0.4 | 95% CI = -0.92, 0.12 Yes | *P*=.003, $I^2$: 79%, No | Yes |
| | N/A | self-harm (frequency) vs control | mean difference: 0.34 | 95% CI: -2.1 - 2.78, No | $I^2$ = 12%, *P*=.39, Yes | Yes |
| **Linardon et al. 2019** | apps | Apps for depression | g= 0.28 | 95% CI: 0.21-0.36 Yes | $I^2$ = 54%, No | Yes (depr sx) |
| | N/A | compared to active controls | g=0.13 | 95% CI: 0.07-0.34, No | $I^2$ = 60%, No | Yes |
| | N/A | with guidance | g=0.48 | 95% CI: 0.34-0.62 No | $I^2$ = 46%, No | Yes |
| | N/A | stand alone | g=0.23 | 95% CI: 0.15-0.31 No | $I^2$ = 32%, Yes | Yes |
| | N/A | Apps for anxiety | g=0.3 | 95% CI: 0.2-0.4, No | $I^2$ = 63%, No | Yes |
| | N/A | compared to active controls | g=0.09 | 95% CI: -0.21-0.39, Yes | $I^2$ = 32%, Yes | Yes |
| | N/A | with guidance | g=0.53 | 95% CI: -0.36-0.70, No | $I^2$ = 60%, No | Yes |
| | N/A | stand alone | g=0.21 | 95% CI: -0.12-0.30, No | $I^2$ = 36%, Yes | Yes |
| | N/A | Apps for social anxiety | g=0.58 | 95% CI: 0.25-0.90) No | $I^2$: 78%, No | Yes |
| | N/A | Apps for panic | g=-0.05 | 95% CI: -0.41- 0.31, No | $I^2$: 0% Yes | Yes |
| | N/A | Apps for PTSD | g=0.18 | 95% CI: -0.04-0.41 - No | $I^2$: 0% Yes | Yes |
| | N/A | outcome: general distress | g=0.40 | 95% CI: 0.24-0.56, Yes | $I^2$: 60% No | No |
| | N/A | outcome: stress | g=0.35 | 95% CI: 0.21-0.48, No | $I^2$ = 62%, No | No |
| | N/A | outcome: quality of life | g=0.35 | 95% CI: 0.29-0.42, Yes | $I^2$ = 24%,Yes | No |

[a]mHealth: mobile health.

[b]N/A: Not applicable.

[c]RCT: Randomized Controlled Trial.

[d]eHealth: electronic health.

**Table 2.** Grade review continued.

| Authors | Study Design | N (total number of participants) | Length of follow-up | Other biases considered | Publication bias considered | Conclusion on effect | Overall Quality |
|---|---|---|---|---|---|---|---|
| | RCT[a], PC (prospective cohort) | P. eg.: 154, controls 145 | No follow-up described | No, but they looked if the effect vary by moderator (supported by a mental health professional) | P. eg. : Yes, small - Yes, No change - No publication bias - unknown | | |
| **Lindhiem et al., 2015** | RCT | 943 | No follow-up described | No | Yes: No change but not specific for apps | medium | poor-moderate |
| **Versluis et al., 2016** | PC | 1008 | No follow-up described | Yes (age, gender, design, etc.) | Yes, No change | medium | poor-moderate |
| | RCT | 481 | No follow-up described | No | Yes, significant risk for bias. Corrected effect size: g=0.23, 95% CI: 0.04-0.42 (considerably smaller) | medium | poor |
| | PC +RCT | 468 | No follow-up described | No | unknown for this subgroup of studies | small-medium | poor |
| | PC | 870 | No follow-up described | No | unknown for this subgroup of studies | small-medium | poor |
| | PC | 199 | No follow-up described | No | unknown for this subgroup of studies | small-medium | poor |
| | PC | 1156 | No follow-up described | No | unknown for this subgroup of studies | small-medium | poor-moderate |
| | PC+RCT | 474 | No follow-up described | No | Unknown for this subanalysis | medium-large | moderate |
| | PC+RCT | 425 | No follow-up described | No | Unknown for this subanalysis | medium | poor |
| **Stratton et al., 2017** | RCT | 2399, controls 2265 (total but only 3 app studies) | follow-up described (g=0.23), but not specific for apps | No | Yes, adjusted effect size = g=0.12, 95% CI = 0.01-0.25 | small | poor-moderate |
| **Firth et al., 2017a** | RCT | 960 (intervention conditions), 877 (control conditions) | No follow-ups described | No | No bias | small-medium | moderate-high |
| | RCT | total: 1026 | No follow-ups described | No | No bias | small | high |
| | RCT | total:1212 | No follow-ups described | No | No bias | small-medium | moderate-high |
| **Firth et al., 2017b** | RCT | 1716, controls 1698 | No follow-up described | Yes (CBT[b]/Not, mindfulness, feedback, person feedback, etc) | No bias *P*=.26 | small-medium | high |
| | RCT | 1195, controls 1186 | No follow-up described | Yes | No bias *P*=.34 | small | moderate-high |
| | RCT | 891, controls 783 | No follow-up described | Yes | No bias *P*=.25 | medium | moderate-high |

| Authors | Study Design | N (total number of participants) | Length of follow-up | Other biases considered | Publication bias considered | Conclusion on effect | Overall Quality |
|---|---|---|---|---|---|---|---|
| | RCT[a], PC (prospective cohort) | P. eg.: 154, controls 145 | No follow-up described | No, but they looked if the effect vary by moderator (supported by a mental health professional) | P. eg. : Yes, small - Yes, No change - No publication bias - unknown | | |
| **Witt et al., 2017** | RCT | 232, controls 236 | 2 studies: -0.34 (CI: -0.70-0.01) | No | unknown | small | moderate |
| | PC | 149 | No follow-up described | No | unknown | small-medium | poor |
| | RCT | 104, controls: 121 | No follow-up described | No | unknown | N.S.[c] | poor-moderate |
| **Linardon et al. 2019** | RCT | 3639, controls: 3519 | 2-6 weeks (n=33): g=0.17, 7-11 weeks (n=18): g=0.46, 12+ (n=3): g=0.09 | Yes, multiple sub-analyses | Yes, results improved when controlled (g=0.41) | small-medium | high |
| | RCT | 526, controls: 530 | | Yes, multiple sub-analyses | | small | moderate-high |
| | RCT | 978, controls: 918 | | | | medium | poor-moderate |
| | RCT | 2489, controls: 2522 | | | | small | moderate |
| | RCT | 2219, controls: 2256 | 2-6 weeks (n=24): g=0.11, 7-11 weeks (n=15): g=0.52, 12+ (n=0) | Yes, multiple sub-analyses | Yes, No change | small-medium | moderate-high |
| | RCT | 134, controls: 137 | No | Yes, multiple sub-analyses | Yes, no change | N.S. | moderate-high |
| | RCT | 859, controls: 827 | | | | medium | poor-moderate |
| | RCT | 859, controls: 860 | | | | small | moderate |
| | RCT | 520, controls: 326 | No | Yes, low risk of bias | unknown | medium | poor-moderate |
| | RCT | 58, controls: 56 | No | Yes, low risk of bias | unknown | N.S. | poor-moderate |
| | RCT | 145, controls: 147 | No | Yes, low risk of bias | unknown | N.S. | poor-moderate |
| | RCT | 919, controls: 949 | No | Yes, low risk of bias | unknown | small-medium | moderate |
| | RCT | 1574, controls: 1711 | 2-6 weeks (n=19): g=0.18, 7-11 weeks (n=6): g=0.63, 12+ (n=2): g=0.59 | Yes, multiple sub-analyses | Yes, results improved when controlled (g=0.44) | small-medium | moderate-high |

| Authors | Study Design | N (total number of participants) | Length of follow-up | Other biases considered | Publication bias considered | Conclusion on effect | Overall Quality |
|---|---|---|---|---|---|---|---|
| | RCT[a], PC (prospective cohort) | P. eg.: 154, controls 145 | No follow-up described | No, but they looked if the effect vary by moderator (supported by a mental health professional) | P. eg. : Yes, small - Yes, No change - No publication bias - unknown | | |
| | RCT | 2714, controls: 2871 | 2-6 weeks (n=31): g=0.35, 7-11 weeks (n=11): g=0.36, 12+ (n=1): g=0.31 | Yes, multiple subanalyses | Yes, results improved when controlled (g=0.39) | small-medium | high |

[a]RCT: Randomized Controlled Trial.

[b]CBT: Cognitive Behavioral Therapy.

[c]N.S.:Not significant.

# Methods

## Literature Search

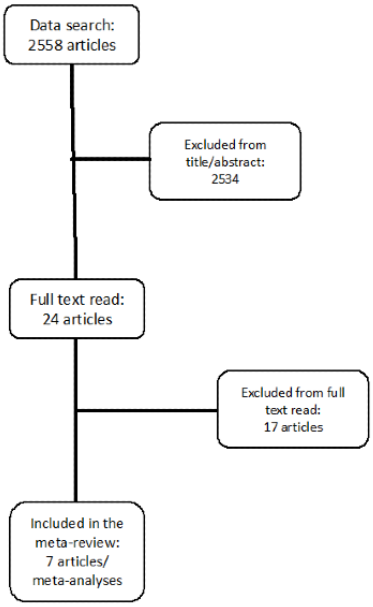
We only included systematic reviews reporting quantitative pooled data (ie, meta-analyses), published in full text, in English or French, and those mentioned the use of app technology for mental health issues.

When more than one meta-analysis was found for a mental health problem, we reviewed them all and used the following criteria to select the ones we kept: (1) if most of the same studies were reviewed, we kept the meta-analysis with the largest number of studies; and (2) between an older meta-analysis with many small uncontrolled studies and a more recent meta-analysis including only RCTs, we chose the latter. We excluded systematic reviews without quantifiable data (eg, qualitative) and treatment guidelines. The final decision to include or exclude reviews was made by consensus by 2 researchers (TL and SP).

## Search Strategy

MEDLINE, EMBASE, Current Contents, PsycINFO, and Google Scholar were searched. Keywords included *mental health*, *technology*, *app*, *mHealth*, *eHealth*, *mobile*, with the added filters: *review* or *meta*. See Figure 1 for the selection of studies.

**Figure 1.** Flow diagram.




## Grading of Recommendations, Assessment, Development, and Evaluation System

The GRADE system was used to assess evidence quality [42]. According to this assessment system, the quality of evidence of meta-analyses can be judged based on various factors, namely, the size of the sample (the larger the better, ideally over 1000), the precision of effects (ie, the CI is not too wide; we opted for within 25% higher or lower than the effect size as ideal), the directness of the outcomes (eg, impact on mental health symptoms [direct] vs impact on perceived stress [indirect]), homogeneity of effects across studies (ie, consistency of results from one study to the next), the study design (prospective studies or RCTs obtain higher scores than

cross-sectional or retrospective studies), follow-up data (if any and the length of time), and publication bias (if analyzed and presented). We also added a specific section for the confounding factors considered, which can include controlling for biases, trial quality, and other variables that could influence the results. The magnitude of the impact of the app is determined based on the estimated effect size (the larger the value, the better) [43]. We chose to present effect sizes apart from the quality of evidence for each study. As such, a point is given for each element of the GRADE system measured, with meta-analyses being rated as either very poor, poor, poor to moderate, moderate, moderate to high, high, or very high–quality evidence. No points are given for the effect size. For each meta-analysis, 2 expert raters (TL and SP) rated the different components with the GRADE system. Both raters met and went over their ratings for a final consensus. Given the stringent criteria involved, consensus was easily reached (over 95% initial agreement). For each component of the model, we present both the quality of the evidence and the effect size. Given that meta-analyses also conducted subanalyses, we reported those that compared the apps with a control condition and indicated the results for stand-alone apps versus apps with guidance.

# Results

## Included Studies

Overall, our search retrieved 2558 potential papers. After excluding irrelevant papers and articles that did not respond to our inclusion criteria, we retrieved 24 meta-analyses that were reviewed, of which 7 were included in the meta-review. Please refer to Figure 1 for the flow diagram of the inclusion of meta-analyses in the meta-review.

## Mental Health (Multiple Problems)

Two meta-analyses [44,45] included apps targeting multiple mental health problems, ranging from anxiety, depression, to substance misuse, and even included some studies on physical health problems or stress. The Lindheim et al's study [44] specifically targeted whether apps offered additional benefits to ongoing treatments or psychotherapy. As such, they only included studies that used apps in addition to a regular (in person) delivered intervention. Overall, the effect size was medium, suggesting that apps can add value to existing treatments. However, the quality of the evidence was rated as poor to moderate (see Tables 1 and 2), given that the effects were imprecise, the samples were very heterogeneous, and the effects were indirect (no subanalyses by diagnosis or problem and all mixed together). However, the meta-analysis included only RCTs, with a total sample size slightly below the criterion of 1000 and verified publication bias. The meta-analysis by Versluis et al [45] was interested in EMI as a tool to increase self-management to cope with depression, anxiety, or stress. As can be seen in Tables 1 and 2, they calculated the effect size in general (all mental health problems together) as well as according to specific outcomes. For the results as a whole, the effect size was medium, the sample size was significant (above 1000), and the publication bias verified, but the other elements did not support quality evidence, with a rating of poor to moderate.

Another meta-analysis also targeted mental health as a larger construct, but within the workplace. This meta-analysis, from the study by Stratton et al [46], however, included various electronic health strategies, with only 3 studies specifically offering an app. The results of these studies are aggregated with other results. As a consequence, the results are indirect, highly heterogeneous, and imprecise with a total quality score of moderate. The effect size was small and decreased when publication bias was included. Nonetheless, the study included a large sample (more than 2000 participants) and only looked at RCTs.

## Anxiety

Versluis et al [45], Firth et al [39], and Linardon et al [47] specifically measured the effect sizes of apps for anxiety symptoms. Although Linardon et al's [47] meta-analysis is more recent and includes many more studies than the other two meta-analyses, it does not include all the studies found in the two previous studies but has many others, justifying the need to keep all 3 meta-analyses in this review. In a study by Versluis et al [45], a medium effect size was found for EMI on self-management of anxiety symptoms, with poor-quality evidence (because of sample size, heterogeneity of samples, imprecise effect, and no follow-up). The study did, however, look at publication biases and included both RCTs and prospective studies. Firth et al [39], on the other hand, included only RCTs and compared apps with waitlist or active controls and found a small to medium effect size overall when compared with waitlist and small effect when compared with active controls. The meta-analysis included homogeneous samples, samples more than 1000 participants and, overall, were rated of moderate to high quality (but high-quality evidence for the comparison with active controls and waitlists). Finally, Linardon et al [47], focusing on generalized anxiety disorder symptoms, also only included RCTs and considered publication bias (which increased the effect size), and included various controls (waitlist and different types of control conditions: information, placebo/attention, and active controls) and found a small to medium effect size (for all controls together). A closer examination revealed that the effect size decreased as the control condition became more stringent, with the effect no longer being significant when an active treatment control was used. They also looked at some follow-up data and found that the effect size remained small for follow-ups of 2 to 6 weeks. However, those (15 studies) that included follow-ups at 7 to 11 weeks found a medium effect size (g=0.52; 95% CI 0.41 to 0.63). This meta-analysis also considered various subgroup analyses (type of app, intervention model, and specific techniques), but these did not seem to modify the outcome. Overall, we rated this meta-analysis as being of high quality (overall) and moderate quality when compared with active controls because of the strengths mentioned and the fact that the results were either imprecise or inconsistent (or small N for active controls).

### *Specific Anxiety Symptoms*

Linardon et al [47] also looked at specific anxiety symptoms, namely, social anxiety, panic, and posttraumatic stress symptoms. Only apps focusing on social anxiety (6 studies) reported a significant medium effect size, with quality evidence

of poor to moderate quality (see Tables 1 and 2). Panic and posttraumatic stress symptoms did not improve in the studies reviewed (3 and 4 studies, respectively), with the evidence rated as poor to moderate quality.

### Depression

In total, 3 meta-analyses measured the impact of apps on symptoms of depression and 1 looked at apps for suicidal ideation and self-harm. As was mentioned for anxiety disorders, Linardon et al's [47] meta-analysis is the most recent but does not include all the studies reviewed in either Versluis et al's [45] or Firth et al's [40] meta-analyses, justifying the need to keep all 3 in this meta-review. Versluis et al [45], looking at EMI for self-management of depressive symptoms, found a small to medium effect size, but the quality of the evidence was judged as poor, given the heterogeneity of the samples, the imprecise effect, the study design (no RCTs), the sample size, and the absence of follow-up. The effect was direct, and publication biases were considered. Firth et al [40] compared smartphone interventions with active and inactive controls and only included RCTs. The overall quality of this study was rated as high, with small to medium effect size overall, medium effect size with inactive controls, and small for active controls. Apart from the inconsistency (heterogeneity) and absence of follow-ups, all other quality criteria were met. As for Linardon et al [47], the effect size was small to medium, the effect precise, all studies included were RCTs, a large sample, with no negative effect of publication bias (in fact an increase was noted). Furthermore, follow-ups were reported for some studies, indicating that the effect size was small at posttreatment and at 2 to 6 weeks, but medium at 7 to 11 weeks follow-up (g=0.46, 95% CI 0.36-0.55). The quality of the evidence was also rated as high (overall), given that heterogeneity was found.

Witt et al [48] conducted a meta-analysis on the use of apps for the self-management of suicidal ideation and self-harm. The apps included were solely stand alone. They conducted analyses of suicidal ideation scores, suicidal behaviors, and self-harm behaviors. As can be seen in Tables 1 and 2, when only including RCTs for suicidal ideation, the effect size was small, imprecise, but the sample was homogeneous, followed up with a similar effect size, and the effect was direct. The quality of the evidence was rated as moderate, given the small sample size and the lack of control for biases (publication or otherwise). When looking at noncontrolled studies for suicidal ideation, the quality of the evidence drops to very poor, with small sample size, high heterogeneity, and imprecise effect. As for self-harm, the analyses were mean differences in the frequency of behavior, with nonsignificant effect and poor to moderate–quality evidence.

### Other Mental Health Concepts

Versluis et al [45] also measured the impact of EMI apps on perceived stress, quality of life, acceptance, and relaxation. We chose to only consider perceived stress and quality of life, given that the latter two are theory or intervention specific. Both had small to medium effect sizes with poor quality for perceived stress and poor to moderate–quality evidence for the quality of life (only precision and sample size offered a point). Linardon et al [47] also included indirect measures, namely, distress, stress, and quality of life. They reported a small-medium effect size, but overall, moderate-quality evidence for distress. For stress, the effect size was small to medium but with moderate to high–quality evidence (thanks to various biases controlled for, follow-up data, large sample, and including only RCTs). As for the quality of life, the effect was also small to medium, but the quality of the evidence was high (thanks to precise, consistent effect, large sample, follow-up, and biases controlled for).

Regarding stand-alone apps versus apps offered with guidance or adjunctive to therapy, only some meta-analyses actually compared these, whereas other meta-analyses looked at only one condition. As such, Lindheim et al's meta-analysis [44] only included adjunctive and had a medium effect size, with poor to moderate quality. Witt et al [48] only included stand-alone apps and found a small effect size, with moderate-quality evidence. Versluis [45] found a medium to large effect size when guidance was offered, compared with medium effect size for stand-alone apps, with stand-alone apps being supported by poor evidence compared with poor to moderate evidence for guidance. Linardon et al's meta-analysis [47] broke down the guidance versus stand-alone apps according to symptoms targeted (ie, anxiety or depression). For anxiety, the effect size is medium for guidance, compared with small for stand-alone apps, with the quality of the evidence being moderate to high for stand-alone and moderate for apps with guidance. For depression, the effect size was medium for apps with guidance versus small for stand-alone apps, with the quality of the evidence being moderate to high for stand-alone apps and moderate for guidance.

## Discussion

### Principal Findings

This meta-review allowed us to closely examine the quality of the evidence reported by 7 meta-analyses (including various subanalyses) on the use of apps for mental health issues. The results are equivocal, with 14 results being linked to poor or poor to moderate, 15 to moderate or moderate to high, and 8 to high-quality evidence.

When examining studies that include various types of apps for mental health (general), we find that the conclusions are not solid with poor to moderate or moderate–quality evidence, although medium effects (or small effects when looking at work) are reported. For higher quality evidence, samples need to be larger, more homogeneous, with biases and follow-ups included. Although it might be tempting to conduct these larger analyses by merging various apps focusing on different mental health issues, they might not convey quality evidence that is useful.

### Specific Findings

Apps for anxiety symptoms appear to bring a clear benefit of small to medium amplitude, but with good-quality evidence. There are some discrepancies in the results reported, with Firth et al [39] seeing a small effect size when apps were compared with active controls, but in a study by Linardon et al [47], a significant effect was not observed when active controls were used for comparison. This might be because of the inclusion

criteria used in these meta-analyses (generalized anxiety symptoms vs anxiety symptoms) or to the much larger sample included in Firth's analysis. Although follow-ups have only been conducted in a limited number of studies, these report sustained benefits at 6 to 11 weeks. Given that we do not know the frequency of people actually using the apps in the studies (daily, weekly, or less), these results are very promising. The results for specific anxiety problems are of lower quality evidence and did not report a significant clinical effect (for posttraumatic stress disorder or panic disorder), except for social anxiety disorder, which is supported by moderate-quality evidence and a medium effect size. Of import, very few studies focused on apps for specific anxiety disorders.

When looking at apps focusing on depressive symptoms, we obtained small to medium effect sizes compared with waitlist, small when compared with active controls, with overall good-quality evidence (especially for more recent meta-analyses). Furthermore, studies reporting follow-ups show maintenance of the effect at 7 to 11 weeks. These results support the use of apps for depression. The quality of the evidence at this time moderately supports apps for suicidal ideation, with a small effect size but does not support apps for self-harm (no effect).

As for indirect mental health outcomes, namely, outcomes that were considered but were not the main focus of the app intervention (such as distress, stress, or quality of life), the effects are consistently small to medium, with greater quality evidence for the most recent meta-analysis [47].

## Limitations

Our results are limited by its focus on mental health. As such, we did not consider apps that focused on a specific intervention or model (eg, mindfulness apps or CBT) and that did not include symptoms as an outcome. Our results also need to consider what we were not able to measure. Although we sought meta-analyses pertaining to apps in mental health, we did not find meta-analyses for multiple domains or mental health problems for which apps have been developed (eg, severe mental illness, addiction, eating disorders, and obsessive-compulsive disorder). Furthermore, few of the reviews considered confounding factors, such as the actual frequency or time of exposure to the app. Although Weisel et al [41] do not recommend stand-alone apps for mental health problems, our results are more nuanced. Indeed, effect sizes tend to be higher for apps that are used with guidance or with an ongoing treatment (medium effect) compared with stand-alone (small effect), but the quality evidence is better for stand-alone apps. This suggests that stand-alone apps mostly offer a small improvement, but this improvement is consistent across quality studies. As such, apps could be used as a stand-alone treatment, while being on a waitlist for an active treatment, for instance, and offer a small effect on symptoms or offered with some guidance or alongside an ongoing in-person treatment for a medium effect on symptoms.

Furthermore, various studies did not use similar control conditions. The use of different types of control group usually leads to variations in effect estimates. The effect sizes of interventions are typically lower when compared with active controls instead of inactive controls [49,50]. We cannot exclude a digital-placebo effect related to the use of the device itself or from the expectations' effect [51] rather than from possible active components [52]. Several recent protocols include a placebo intervention (a sham version of the app) [53]; unfortunately, only some of the studies assessed in the included meta-review involved such placebo app control. Furthermore, several studies were conducted with nonclinical populations, who presented with symptoms but perhaps not a diagnosed disorder, limiting the generalizability of the results for clinical populations [54].

Nonetheless, the nature of smartphone interventions does appear to position them as a possible low-intensity intervention tool for those with less severe levels of symptoms or as a first step in a stepped-care approach to service delivery [55]. The follow-up data available to date also suggest that gains are sustainable over a few months. Additional follow-up data are warranted to confirm these results.

Attrition is another problem repeatedly described in smartphone app–related studies [56] and in naturalistic use [57]. Further studies should include a detailed description of the behavior change techniques involved in the design [58] as well as data on the actual utilization of the different app functions. It will be helpful to increase our knowledge about effective strategies in behavior change as well as about the app use engagement. It would also be useful to have a better understanding of the context in which the app is used, at home, at work, at the clinic in the waiting room, alone, or with a therapist or a family member.

## Conclusions

We believe that future studies should focus on high users of apps, namely, youth and young adults. We currently do not have specific information on the efficacy and actual use of mental health apps with such subgroups of individuals. To date, most of the app studies on mental health have focused on feasibility and acceptability, with only a small portion actually pushing forward toward efficacy trials (and often with small numbers). The field of apps for mental health is burgeoning, with the speed of delivery of the app being a primary concern. Traditional study designs (such as RCTs) tend to take a long duration to complete and can deter app developers who aim to commercialize their product. Other controlled research designs could be encouraged (eg, repeated single-case experimental designs) to encourage quality studies at a more rapid speed.

In conclusion, apps for anxiety and depression hold great promise with clear clinical advantages, modestly as stand-alone self-management, and more strongly with guidance or adjunctive treatments. More meta-analyses and more quality studies are needed to recommend apps for other mental health issues or for specific populations.

## Conflicts of Interest

None declared.

## References


1. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
2. Bradway M, Carrion C, Vallespin B, Saadatfard O, Puigdomènech E, Espallargues M, et al. mHealth assessment: conceptualization of a global framework. JMIR Mhealth Uhealth 2017 May 2;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
3. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120 [FREE Full text] [doi: 10.2196/jmir.2600] [Medline: 23770578]
4. van Singer M, Chatton A, Khazaal Y. Quality of smartphone apps related to panic disorder. Front Psychiatry 2015;6:96 [FREE Full text] [doi: 10.3389/fpsyt.2015.00096] [Medline: 26236242]
5. Penzenstadler L, Chatton A, van Singer M, Khazaal Y. Quality of smartphone apps related to alcohol use disorder. Eur Addict Res 2016;22(6):329-338. [doi: 10.1159/000449097] [Medline: 27598779]
6. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
7. Choi J, Noh G, Park D. Smoking cessation apps for smartphones: content analysis with the self-determination theory. J Med Internet Res 2014 Feb 12;16(2):e44 [FREE Full text] [doi: 10.2196/jmir.3061] [Medline: 24521881]
8. Azar KM, Lesser LI, Laing BY, Stephens J, Aurora MS, Burke LE, et al. Mobile applications for weight management: theory-based content analysis. Am J Prev Med 2013 Nov;45(5):583-589. [doi: 10.1016/j.amepre.2013.07.005] [Medline: 24139771]
9. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological momentary assessment and intervention in the treatment of psychotic disorders: a systematic review. Psychiatr Serv 2017 Nov 1;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
10. World Health Organization. 2017. Frequently Asked Questions on Global Task Force on Digital Health for TB and Its Work URL: https://www.who.int/tb/areas-of-work/digital-health/faq/en/ [accessed 2020-03-22]
11. Rahmani AM, Gia TN, Negash B, Anzanpour A, Azimi I, Jiang M, et al. Exploiting smart e-Health gateways at the edge of healthcare Internet-of-Things: A fog computing approach. Future Gener Comput Syst 2018 Jan;78:641-658. [doi: 10.1016/j.future.2017.02.014]
12. International Telecommunications Union. 2016. ICT Facts and Figures - 2016 URL: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2016.pdf
13. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
14. Monney G, Penzenstadler L, Dupraz O, Etter J, Khazaal Y. mHealth app for cannabis users: satisfaction and perceived usefulness. Front Psychiatry 2015;6:120 [FREE Full text] [doi: 10.3389/fpsyt.2015.00120] [Medline: 26379561]
15. Kiluk BD, Carroll KM. New developments in behavioral treatments for substance use disorders. Curr Psychiatry Rep 2013 Dec;15(12):420 [FREE Full text] [doi: 10.1007/s11920-013-0420-1] [Medline: 24173656]
16. Sort A, Khazaal Y. Six tips on how to bring epic wins to health care. Front Psychiatry 2017;8:264 [FREE Full text] [doi: 10.3389/fpsyt.2017.00264] [Medline: 29249994]
17. Torous J, Hsin H. Empowering the digital therapeutic relationship: virtual clinics for digital health interventions. NPJ Digit Med 2018;1:16 [FREE Full text] [doi: 10.1038/s41746-018-0028-2] [Medline: 31304301]
18. Fatseas M, Serre F, Swendsen J, Auriacombe M. Effects of anxiety and mood disorders on craving and substance use among patients with substance use disorder: an ecological momentary assessment study. Drug Alcohol Depend 2018 Jun 1;187:242-248. [doi: 10.1016/j.drugalcdep.2018.03.008] [Medline: 29684892]
19. Jones A, Tiplady B, Houben K, Nederkoorn C, Field M. Do daily fluctuations in inhibitory control predict alcohol consumption? An ecological momentary assessment study. Psychopharmacology (Berl) 2018 May;235(5):1487-1496 [FREE Full text] [doi: 10.1007/s00213-018-4860-5] [Medline: 29497782]
20. Merikangas KR, Swendsen J, Hickie IB, Cui L, Shou H, Merikangas AK, et al. Real-time mobile monitoring of the dynamic associations among motor activity, energy, mood, and sleep in adults with bipolar disorder. JAMA Psychiatry 2019 Feb 1;76(2):190-198 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3546] [Medline: 30540352]
21. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]

22. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
23. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Apr 27;4(2):73-85 [FREE Full text] [doi: 10.1007/s41347-019-00095-w]
24. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002]
25. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J Med Internet Res 2017 Nov 3;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
26. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
27. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of 'mHealth' among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
28. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
29. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
30. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl) 2017;10:237-251 [FREE Full text] [doi: 10.2147/MDER.S144158] [Medline: 29042823]
31. Velligan D, Mintz J, Maples N, Xueying L, Gajewski S, Carr H, et al. A randomized trial comparing in person and electronic interventions for improving adherence to oral medications in schizophrenia. Schizophr Bull 2013 Sep;39(5):999-1007 [FREE Full text] [doi: 10.1093/schbul/sbs116] [Medline: 23086987]
32. Ben-Zeev D, Kaiser SM, Brenner CJ, Begale M, Duffecy J, Mohr DC. Development and usability testing of FOCUS: a smartphone system for self-management of schizophrenia. Psychiatr Rehabil J 2013 Dec;36(4):289-296 [FREE Full text] [doi: 10.1037/prj0000019] [Medline: 24015913]
33. Khazaal Y, Monney G, Richter F, Achab S. Game-control, rational of an application to support the limits of games. Article in French. Jeu-contrôle, rationnel d'une application de soutien aux limites de jeux. J Ther Comport Cogn 2017 Sep;27(3):129-137. [doi: 10.1016/j.jtcc.2017.05.003]
34. Bertholet N, Godinho A, Cunningham JA. Smartphone application for unhealthy alcohol use: pilot randomized controlled trial in the general population. Drug Alcohol Depend 2019 Feb 1;195:101-105. [doi: 10.1016/j.drugalcdep.2018.12.002] [Medline: 30611977]
35. Tait RJ, Kirkman JJ, Schaub MP. A participatory health promotion mobile app addressing alcohol use problems (the Daybreak Program): protocol for a randomized controlled trial. JMIR Res Protoc 2018 May 31;7(5):e148 [FREE Full text] [doi: 10.2196/resprot.9982] [Medline: 29853435]
36. Gustafson DH, McTavish FM, Chih M, Atwood AK, Johnson RA, Boyle MG, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
37. Wright CJ, Dietze PM, Agius PA, Kuntsche E, Livingston M, Black OC, et al. Mobile phone-based ecological momentary intervention to reduce young adults' alcohol use in the event: a three-armed randomized controlled trial. JMIR Mhealth Uhealth 2018 Jul 20;6(7):e149 [FREE Full text] [doi: 10.2196/mhealth.9324] [Medline: 30030211]
38. Crane D, Garnett C, Michie S, West R, Brown J. A smartphone app to reduce excessive alcohol consumption: identifying the effectiveness of intervention components in a factorial randomised control trial. Sci Rep 2018 Mar 12;8(1):4384 [FREE Full text] [doi: 10.1038/s41598-018-22420-8] [Medline: 29531280]
39. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
40. 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;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
41. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]

42. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. Br Med J 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
43. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [doi: 10.1016/j.jclinepi.2010.07.015] [Medline: 21208779]
44. Lindhiem O, Bennett CB, Rosen D, Silk J. Mobile technology boosts the effectiveness of psychotherapy and behavioral interventions: a meta-analysis. Behav Modif 2015 Nov;39(6):785-804 [FREE Full text] [doi: 10.1177/0145445515595198] [Medline: 26187164]
45. Versluis A, Verkuil B, Spinhoven P, van der Ploeg MM, Brosschot JF. Changing mental health and positive psychological well-being using ecological momentary interventions: a systematic review and meta-analysis. J Med Internet Res 2016 Jun 27;18(6):e152 [FREE Full text] [doi: 10.2196/jmir.5642] [Medline: 27349305]
46. Stratton E, Lampit A, Choi I, Calvo RA, Harvey SB, Glozier N. Effectiveness of eHealth interventions for reducing mental health conditions in employees: a systematic review and meta-analysis. PLoS One 2017;12(12):e0189904 [FREE Full text] [doi: 10.1371/journal.pone.0189904] [Medline: 29267334]
47. 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;18(3):325-336 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
48. Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 15;17(1):297 [FREE Full text] [doi: 10.1186/s12888-017-1458-0] [Medline: 28810841]
49. Opriş D, Pintea S, García-Palacios A, Botella C, Szamosközi S, David D. Virtual reality exposure therapy in anxiety disorders: a quantitative meta-analysis. Depress Anxiety 2012 Feb;29(2):85-93. [doi: 10.1002/da.20910] [Medline: 22065564]
50. Pinquart M, Duberstein PR, Lyness JM. Effects of psychotherapy and other behavioral interventions on clinically depressed older adults: a meta-analysis. Aging Ment Health 2007 Nov;11(6):645-657. [doi: 10.1080/13607860701529635] [Medline: 18074252]
51. Gruszka P, Burger C, Jensen MP. Optimizing expectations via mobile apps: a new approach for examining and enhancing placebo effects. Front Psychiatry 2019;10:365 [FREE Full text] [doi: 10.3389/fpsyt.2019.00365] [Medline: 31214057]
52. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. Lancet Psychiatry 2016 Feb;3(2):100-102. [doi: 10.1016/S2215-0366(15)00565-9] [Medline: 26851322]
53. Giosan C, Cobeanu O, Mogoaşe C, Szentagotai A, Mureşan V, Boian R. Reducing depressive symptomatology with a smartphone app: study protocol for a randomized, placebo-controlled trial. Trials 2017 May 12;18(1):215 [FREE Full text] [doi: 10.1186/s13063-017-1960-1] [Medline: 28494802]
54. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
55. Marzano L, Bardill A, Fields B, Herd K, Veale D, Grey N, et al. The application of mHealth to mental health: opportunities and challenges. Lancet Psychiatry 2015 Oct;2(10):942-948. [doi: 10.1016/S2215-0366(15)00268-0] [Medline: 26462228]
56. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
57. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
58. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]


## Abbreviations

**EMI:** Ecological Momentary Interventions
**GRADE:** grading of recommendations, assessment, development, and evaluation
**mHealth:** mobile health
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 13.12.19; peer-reviewed by J Linardon, J Kelland, I Savolainen; comments to author 11.01.20; revised version received 11.02.20; accepted 26.02.20; published 29.05.20.*

*Please cite as:*
*Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, Francoeur A, Pennou A, Khazaal Y*
*Mobile Apps for Mental Health Issues: Meta-Review of Meta-Analyses*
*JMIR Mhealth Uhealth 2020;8(5):e17458*
*URL: https://mhealth.jmir.org/2020/5/e17458*
*doi:10.2196/17458*
*PMID:32348289*

Original Paper

# Adaptive Mobile Health Intervention for Adolescents with Asthma: Iterative User-Centered Development

David A Fedele[1*], PhD, ABPP; Christopher C Cushing[2*], PhD; Natalie Koskela-Staples[1], BA; Susana R Patton[3], PhD, ABPP; Elizabeth L McQuaid[4], PhD, ABPP; Joshua M Smyth[5], PhD; Sreekala Prabhakaran[6], MD; Selina Gierer[7], DO; Arthur M Nezu[8], PhD, ABPP

[1]Department of Clinical & Health Psychology, University of Florida, Gainesville, FL, United States
[2]Clinical Child Psychology Program & Schiefelbusch Institute for Life Span Studies, University of Kansas, Lawrence, KS, United States
[3]Nemours Center for Healthcare Delivery Science, Nemours Children's Health System, Jacksonville, FL, United States
[4]Department of Psychiatry and Human Behavior, Brown University, Providence, RI, United States
[5]Department of Biobehavioral Health, Pennsylvania State University, University Park, PA, United States
[6]Department of Pediatrics, University of Florida, Gainesville, FL, United States
[7]Department of Pediatrics, University of Kansas Medical Center, Kansas City, MO, United States
[8]Department of Psychology, Drexel University, Philadelphia, PA, United States
[*]these authors contributed equally

**Corresponding Author:**
David A Fedele, PhD, ABPP
Department of Clinical & Health Psychology
University of Florida
101 S Newell Dr, Rm 3151
PO Box 100165
Gainesville, FL, 32610
United States
Phone: 1 3522945765
Fax: 1 3522736156
Email: dfedele@phhp.ufl.edu

## Abstract

**Background:** Adolescents diagnosed with persistent asthma commonly take less than 50% of their prescribed inhaled corticosteroids (ICS), placing them at risk for asthma-related morbidity. Adolescents' difficulties with adherence occur in the context of normative developmental changes (eg, increased responsibility for disease management) and rely upon still developing self-regulation and problem-solving skills that are integral for asthma self-management. We developed an adaptive mobile health system, Responsive Asthma Care for Teens (ReACT), that facilitates self-regulation and problem-solving skills during times when adolescents' objectively measured ICS adherence data indicate suboptimal rates of medication use.

**Objective:** The current paper describes our user-centered and evidence-based design process in developing ReACT. We explain how we leveraged a combination of individual interviews, national crowdsourced feedback, and an advisory board comprised of target users to develop the intervention content.

**Methods:** We developed ReACT over a 15-month period using one-on-one interviews with target ReACT users (n=20), national crowdsourcing (n=257), and an advisory board (n=4) to refine content. Participants included 13-17–year-olds with asthma and their caregivers. A total of 280 adolescents and their caregivers participated in at least one stage of ReACT development.

**Results:** Consistent with self-regulation theory, adolescents identified a variety of salient intrapersonal (eg, forgetfulness, mood) and external (eg, changes in routine) barriers to ICS use during individual interviews. Adolescents viewed the majority of ReACT intervention content (514/555 messages, 93%) favorably during the crowdsourcing phase, and the advisory board helped to refine the content that did not receive favorable feedback during crowdsourcing. Additionally, the advisory board provided suggestions for improving additional components of ReACT (eg, videos, message flow).

**Conclusions:** ReACT involved stakeholders via qualitative approaches and crowdsourcing throughout the creation and refinement of intervention content. The feedback we received from participants largely supported ReACT's emphasis on providing adaptive

and personalized intervention content to facilitate self-regulation and problem-solving skills, and the research team successfully completed the recommended refinements to the intervention content during the iterative development process.

***(JMIR Mhealth Uhealth 2020;8(5):e18400)*** doi:10.2196/18400

**KEYWORDS**

asthma; mobile health; adherence; adolescence; self-regulation; problem-solving; adolescent; youth

## Introduction

### Background

Over 8% of youth have an asthma diagnosis, making it the most prevalent pediatric chronic illness [1]. Asthma is a leading cause of emergency department visits, missed school days, and healthcare expenditures, making it a significant public health concern [2,3]. Youth can mitigate asthma-related morbidity via consistent engagement in a complex set of daily disease self-management behaviors (eg, monitoring symptoms, avoiding triggers). According to national guidelines, adherence to inhaled corticosteroids (ICS), medications designed to control asthma and reduce the likelihood of exacerbations, is critical to asthma self-management for youth with persistent asthma [4]. High levels of adherence to ICS (ie, taking >80% of prescribed doses) are associated with both consistent asthma control in youth [5] and reduction in severe asthma exacerbations [6].

Adolescence is a unique developmental period when suboptimal adherence to ICS is common [7-9]. As adolescents' responsibility for disease management (eg, taking ICS) increases, there is less direct caregiver contact and an increased desire for autonomy [8]. Concurrently, adolescents' executive function abilities that undergird the self-regulation and problem-solving skills integral for successful asthma self-management are not fully developed [10-15]. Together, these factors put adolescents at high risk for asthma-related morbidity and reduced quality of life [16-19]. There are evidence-based face-to-face programs available addressing adolescent adherence; however, there are often numerous barriers to implementation including infrequent encounters at the point of care and logistical obstacles (eg, transportation, absence of trained interventionists) [20]. Thus, there continues to be a large number of youth who do not take their medication as prescribed [21], leading many researchers to pursue novel intervention frameworks for improving adherence among adolescents.

Mobile health (mHealth) interventions have recently received considerable attention for improving adherence to ICS among youth [20,22]. Smartphones are a readily available intervention medium for youth with asthma given their ubiquitous nature across socioeconomic strata and habitual daily use by adolescents [23,24]. Furthermore, recent advances in ambulatory disease management technology (eg, medication sensors) have led to opportunities in mHealth for continuous passive monitoring to better identify and contextualize states of vulnerability for poor or declining disease self-management. These technological advances have led to burgeoning interest in the development of adaptive mHealth interventions where youth could receive in-the-moment support during periods when a person with a disease may be most in need [25]. To our knowledge, no adaptive mHealth systems exist that objectively monitor youth adherence to ICS and use that data to deliver a tailored, timely, and theory-based intervention.

Our interdisciplinary team of behavioral health experts, pediatric pulmonologists, and technologists recently developed an adaptive mHealth adherence promotion intervention, Responsive Asthma Care for Teens (ReACT), that is grounded in self-regulation theory (SRT) [23]. ReACT is an integrated mHealth system that passively monitors adolescents' adherence to ICS using a novel Bluetooth-enabled sensor, developed in collaboration with the University of Kansas Instrumentation Design Laboratory, that attaches to ICS canisters. ReACT activates when an adolescent's ICS adherence data indicate a clinically-derived need (ie, <80% [5]) and delivers tailored intervention content via the Way to Health text messaging platform. Once active, ReACT uses 2 components, a goal-setting algorithm and tailored problem-solving modules, to support adolescents' self-reaction to suboptimal adherence and improve self-efficacy. ReACT's goal-setting algorithm aids adolescents in self-monitoring, feedback, and goal setting via gain-framed [24] messages. ReACT uses algorithms that assess an adolescent's recent patterns of adherence (eg, trajectory, patterns of missing doses) in order to deliver meaningful goal intention–formatted message content. For example, at times when adherence is suboptimal, ReACT delivers a brief motivation assessment. If upon completing this assessment, the adolescent endorses motivation to take at least some of his/her medication, then ReACT will ask the adolescent to report any intrapersonal or external barriers to ICS adherence (eg, stress) so that ReACT can deliver a tailored problem-solving module while also continuing to monitor the adolescent's ICS use and provide adherence feedback and goal-setting content. Alternatively, if an adolescent indicates that he or she is not currently interesting in taking their ICS, then ReACT provides educational content focused on the importance of medication adherence and delays restarting the goal-setting algorithm for 3 days, giving the adolescent time to process ReACT's educational message.

### Current Study

This study illustrates the iterative user-centered design process we used in developing ReACT, which is consistent with best practices for the development of mHealth pediatric intervention content [26]. Specifically, we describe how we leveraged a combination of individual interviews, national crowdsourced feedback, and an advisory board comprised of target users to develop the core ReACT intervention content and supporting technical and device-related infrastructure. All study procedures were approved by the Institutional Review Boards at the University of Florida and University of Kansas.

## Methods

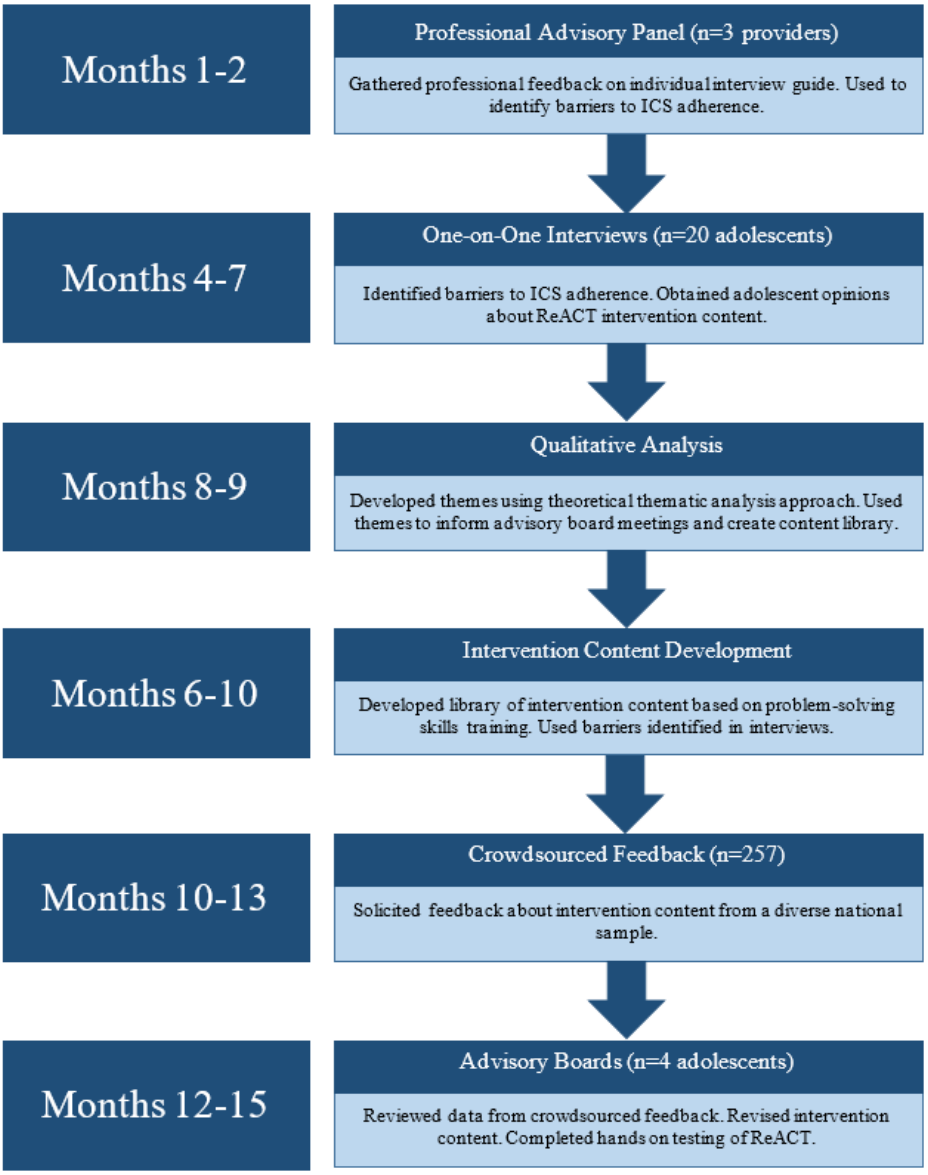
### Study Stages

Our intervention development process consisted of 6 stages, with refinements occurring between iterations. Figure 1 provides an overview of the study procedures and timeline.

### Participants

Participants included 13-17–year-olds with asthma and their caregivers. Individual interviews were completed by 20 adolescent-caregiver dyads (10 from the University of Florida and 10 from the University of Kansas Medical Center), 257 adolescents provided crowdsourced feedback via a national online panel, and 4 adolescent-caregiver dyads participated in advisory board meetings. One dyad participated in both individual interviews and advisory board meetings, yielding a total of 280 adolescents and 23 caregivers. Information regarding recruitment is provided below. For individual interviews and advisory board meetings, dyads were eligible to participate if the adolescent had a physician-verified diagnosis of asthma with persistent symptoms requiring ICS use for at least 6 months, the adolescent had a daily ICS or ICS/long-acting beta agonist prescription for at least 6 months, and the adolescent and caregiver were fluent in English. We excluded dyads from individual interviews and advisory boards if the adolescent had a comorbid chronic health condition that may affect lung function (eg, cystic fibrosis) or if the adolescent had a significant cognitive impairment or developmental delay that could interfere with study completion. For crowdsourced feedback, adolescents were eligible to participate if their caregiver provided affirmative answers to the following questions: “Have you ever been told by a doctor, nurse, or other health professional that your child has asthma?” and “Does your child who is 13-17 still have asthma?” Epidemiological trials (eg, National Health Interview Survey) commonly use these questions to screen for persistent asthma [27].

**Figure 1.** Study overview and timeline. ICS: inhaled corticosteroids; ReACT: Responsive Asthma Care for Teens.



### Procedure

We recruited a convenience sample of individual interview and advisory board meeting participants via clinics and flyers at both sites. At the University of Florida, we also recruited participants via a database of patients that consented to be contacted for research. We used Qualtrics, a leading online panel and survey technology provider, for crowdsourcing to ensure a nationally representative sample of adolescents with asthma in regards to race/ethnicity. Participants completed a screener that ensured that children were between the ages of 13 years and 17 years, had received a diagnosis of asthma, were physically present to participate in the survey, and still had asthma. For participants recruited via Qualtrics, we set the following quotas in order to solicit feedback from youth with asthma from a diverse range of ethnic backgrounds: 30% Black/Non-Hispanic, 24% more than one race, 16% White/Non-Hispanic, 14% Asian, 4% American Indian/Alaska Native, 4% Native Hawaiian or other Pacific Islander, 4% White/Hispanic, and 4% Black/Hispanic. We combined all quotas for non-White participants into one group to allow for faster data collection after 2 months of recruitment. Please see our previous work for more detail about the recruitment procedures [23].

#### *Design Phase I*

Design Phase I consisted of individual interviews with 20 adolescents diagnosed with persistent asthma. The study team, in conjunction with a professional advisory panel comprised of a pediatric pulmonologist, clinical pharmacist, and advanced practice registered nurse, created a semistructured interview guide. The guide assessed salient intrapersonal and external adherence barriers prior to the start of interviews and the types of intervention content adolescents would prefer in an mHealth intervention when encountering these barriers. The guide was informed by SRT [28], our pilot data [29], and the existing pediatric asthma literature on adherence to ICS [30,31]. We used a graduate student (NS) at the University of Florida and a research coordinator at the University of Kansas to conduct the interviews. The principal investigators (DF and CC) trained research staff in interview administration and conducted at least two mock interviews prior to conducting participant interviews. Interviews were 30-60 minutes and audiorecorded. Adolescents and their caregivers completed asthma-related questionnaires [23] in order to describe the sample and gather data on asthma-related and adherence-related constructs. Adolescent-caregiver dyads received US $60 as compensation for their participation.

A transcription service completed verbatim transcriptions of e-recordings of individual interviews. We entered interview files into NVivo (QSR International, Doncaster, Australia), a qualitative data management system. Two research assistants coded and aggregated interviews using a theoretical thematic analysis approach for developing themes [32-34]. This approach used a priori thematic categories guided by SRT, although we allowed for de novo themes to emerge from interviews. We resolved differences via discussion between research assistants and one of the principal investigators (DF). Subsequently, we used the information gathered from individual interviews, preliminary data, and the extant asthma literature to develop intervention content for ReACT.

#### *Design Phase II*

Design Phase II consisted of gathering feedback on ReACT intervention content via national crowdsourcing via Qualtrics and an advisory board of target ReACT users. The 7 different domains of content and branching logic included in ReACT resulted in approximately 80 pieces of content to rate per domain. Thus, we elected to allow each participant to view content from only 1 domain to reduce participant burden. An average of 33 participants (range 28-35 participants) viewed content for each domain and rated its appropriateness using a dichotomous response choice: “yes” (I like the message as it is) or “no” (change it to make it better). When crowdsourcing participants answered “no,” they had the option to reword the message to make it better. Consistent with previous research [35], content receiving ≥60% “no” votes was discarded, and content receiving ≤39% “no” votes was accepted as final content. Content receiving 40%-59% “no” votes was subject to revision. Finally, to complete this stage of content review and refinement, an advisory board comprised of 4 adolescents diagnosed with persistent asthma from the University of Florida convened 3 times over the span of 4 months to refine the intervention content based on feedback from the crowdsourcing. Advisory board members and our previous individual interview participants completed the same asthma-related questionnaires. During advisory board meetings, we audiotaped their comments and transcribed the tapes to inform ReACT design decisions. Adolescent-caregiver dyads received US $50 for each advisory board (total of US $150) as compensation for their participation.

## Results

### Design Phase I

#### *Individual Interviews with Adolescents Diagnosed with Asthma*

We conducted individual interviews with adolescents diagnosed with persistent asthma to identify what intrapersonal and external barriers to adherence to ICS are most salient to adolescents with asthma and to solicit their opinion about the types of intervention content that an mHealth intervention should deliver. We conducted the individual interviews with SRT [28] in mind; we asked about components of SRT if they were not mentioned or probed about how participants’ comments may be related to SRT. Multimedia Appendix 1 presents our individual interview guide questions. Table 1 presents the characteristics of the individual interview and advisory board participants. With regards to diversity, ≥50% of our sample was comprised of racial and ethnic minority groups.

**Table 1.** Youth, caregiver, and family demographic and medical characteristics of the interview and advisory board participants.

| Characteristics | UF[a] (n=13) | KU[b] (n=10) |
|---|---|---|
| Youth age (years), mean (SD) | 15.1 (1.04) | 15.7 (0.95) |
| Caregiver age (years), mean (SD) | 45.5 (10.90) | 44.3 (8.22) |
| **Youth gender, n (%)** | | |
| Female | 7 (54) | 7 (70) |
| Male | 6 (46) | 3 (30) |
| Other | 0 (0) | 0 (0) |
| **Caregiver gender, n (%)** | | |
| Female | 12 (92) | 10 (100) |
| Male | 1 (8) | 0 (0) |
| Other | | |
| **Youth race, n (%)** | | |
| Black/African American | 3 (23) | 0 (0) |
| Caucasian | 6 (46) | 7 (70) |
| Multiracial | 4 (31) | 3 (30) |
| **Youth ethnicity, n (%)** | | |
| Non-Hispanic/Latino | 10 (77) | 6 (60) |
| Hispanic/Latino | 3 (23) | 4 (40) |
| Non-Hispanic Caucasian youth, n (%) | 5 (38) | 5 (50) |
| **Frequency of youth asthma attacks (past year), n (%)** | | |
| A few times a week | 1 (8) | 3 (30) |
| A few times a month | 6 (46) | 4 (40) |
| About once a month or less | 6 (46) | 3 (30) |
| **Youth asthma-related emergency department visits (past 4 weeks), n (%)** | | |
| 0 | 12 (92) | 9 (90) |
| 1 | 1 (8) | 1 (10) |
| **Youth asthma-related emergency department visits (past year)** | | |
| 0 | 6 (46) | 8 (80) |
| 1 | 4 (31) | 1 (10) |
| 2 | 1 (8) | 1 (10) |
| 3 | 1 (8) | 0 (0) |
| 5 | 1 (8) | 0 (0) |
| **Youth asthma-related sick visits, mean (SD)** | | |
| Past 4 weeks | 0.6 (0.9) | 0.3 (0.7) |
| Past year | 2.6 (2.3) | 2.7 (3.7) |
| **Youth school days missed due to asthma, mean (SD)** | | |
| Past 4 weeks | 1.5 (4.1) | 0.4 (1.7) |
| Past year | 9.9 (20.0) | 1.7 (4.7) |
| **Frequency of youth quick relief medication use (past 4 weeks), n (%)** | | |
| Never | 4 (31) | 3 (30) |
| 0-2 days a week | 5 (38) | 4 (40) |
| 3-6 days a week | 1 (8) | 2 (20) |
| Every day of the week | 3 (23) | 1 (10) |

| Characteristics | UF[a] (n=13) | KU[b] (n=10) |
|---|---|---|
| **Caregiver education, n (%)** | | |
| High school | 5 (38) | 4 (40) |
| Some college | 1 (8) | 1 (10) |
| College | 3 (23) | 4 (40) |
| Graduate school | 2 (15) | 0 (0) |
| Other | 2 (15) | 1 (10) |
| **Family income (US $), n (%)** | | |
| <12,000 | 0 (0) | 2 (20) |
| 12,000-24,999 | 3 (23) | 1 (10) |
| 25,000-49,999 | 4 (31) | 3 (30) |
| 50,000-99,999 | 3 (23) | 2 (20) |
| ≥100,000 | 1 (8) | 1 (10) |
| No response | 2 (15) | 1 (10) |

[a]UF: University of Florida.

[b]KU: University of Kansas.

### *Individual Interview Themes*

Table 2 includes the themes of intrapersonal and external barriers endorsed by the adolescents. Adolescents noted a range of intrapersonal barriers to ICS adherence including forgetting (19/20, 95%), difficulties with time management or having a busy schedule (16/20, 80%), and being too fatigued or tired (12/20, 60%). Notably, less than half our sample endorsed a range of additional intrapersonal barriers (eg, mood, stress, laziness). Regarding external barriers to adherence, 12 adolescents (12/20, 60%) reported changes to their routine (eg, being away from home) as barriers. Adolescents also frequently endorsed not having medication available (11/20, 55%) and interference from other activities (10/20, 50%).

Adolescents provided several suggestions about how an mHealth intervention could promote ICS adherence. All adolescents endorsed that an mHealth intervention should provide reminders to take ICS medications and suggested the frequency of notifications from an mHealth intervention should be each time a dose is scheduled (15/20, 75%) or to be flexible dependent upon need (13/20, 65%). They also generally agreed that an mHealth intervention should provide personalized notifications (13/20, 65%) and facilitate tracking of adherence over time (20/20, 100%). Finally, a number of adolescents endorsed that they were receptive to an mHealth intervention sending text messages (10/20, 50%) and including interactive videos (10/20, 50%) to deliver content.

**Table 2.** Adolescent-endorsed intrapersonal and external adherence barriers to inhaled corticosteroid adherence.

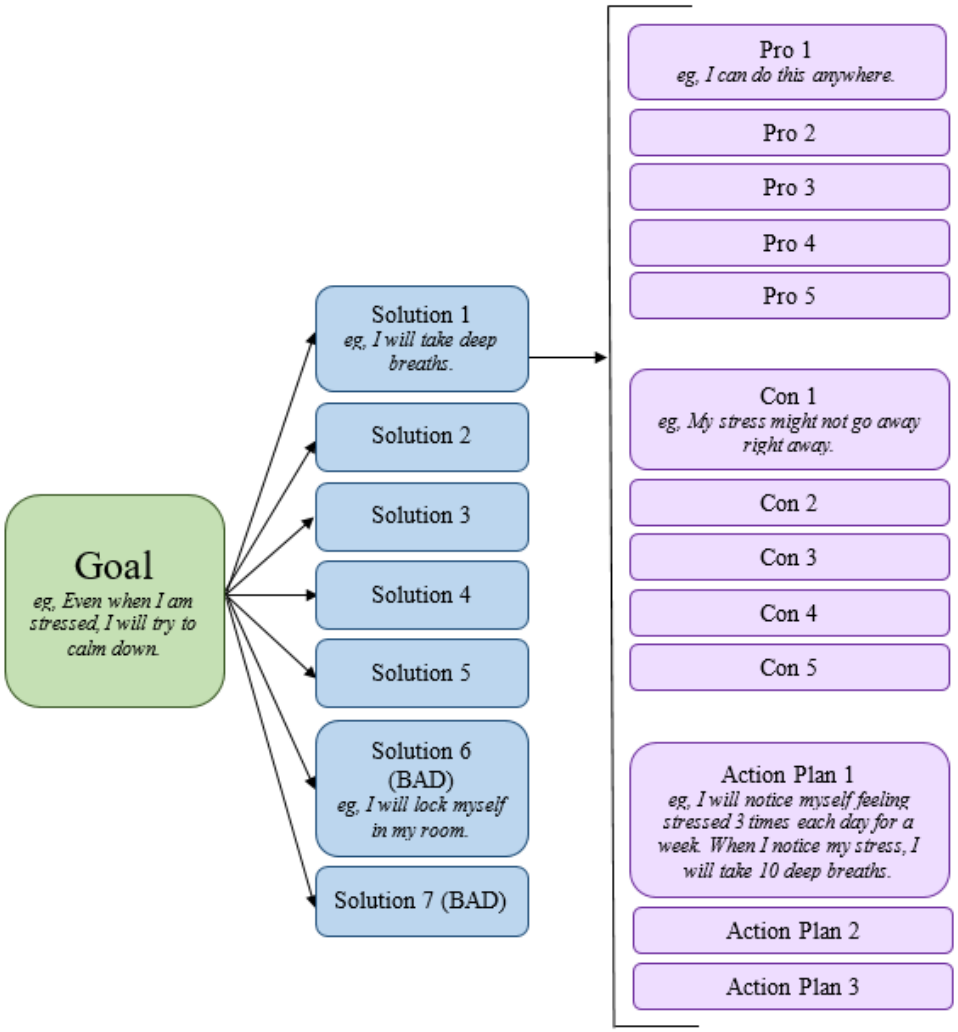
| Interview themes | Participants who endorsed, n (%) | Number of times a barrier was mentioned | Description of barrier | Sample quotes |
|---|---|---|---|---|
| **Intrapersonal barriers** | | | | |
| Forgetting | 19 (95) | 71 | The teen describes forgetting to take their medicine. | "I definitely forget to take it at least 2-3 times a week." |
| Time management/busyness | 16 (80) | 35 | The teen does not take medicine due to busyness or other activities (eg, going to school or work). | "Having to take medicine on a daily basis is a bit difficult for me because I have a busy schedule. I'll be more focused on something and end up forgetting." |
| Sleep/fatigue | 12 (60) | 21 | The teen does not take medicine due to abnormal sleep patterns, difficulty waking up in the morning, going to sleep late, or being fatigued. | "If I'm really tired or I had a long day or I got back from a soccer game, usually if it's a late night, I'll forget to take it because I'll be so tired." |
| Mood | 8 (40) | 11 | The teen's mood (eg, depression, anxiety, sadness, frustration, anger) is a barrier that prevents them from taking medicine. | "If I'm kind of like in a 'eh' mood, I won't take them, but if I'm happy, I'll take them. Yeah, it definitely depends on my mood." |
| Not wanting to take meds | 6 (30) | 12 | The teen mentions not wanting to, deciding not to, or not feeling like taking meds without giving a reason that would better fit in another category. | "Besides forgetfulness, sometimes I just don't feel like taking it." |
| Stress | 4 (20) | 5 | The teen describes stress as a barrier to taking medicine. | "I guess sometimes if I'm super tired or stressed, then I won't really focus on [taking meds]." |
| Embarrassment | 3 (15) | 5 | The teen describes embarrassment as a barrier to taking medicine (eg, being embarrassed to take medicine in front of friends). | "It's kind of embarrassing to pull out an inhaler right before games… It's just not being embarrassed to take it in front of other people I guess." |
| Laziness | 3 (15) | 3 | The teen mentions laziness as a barrier to taking medicine. | "It's just my own, I don't know, laziness I guess that makes it harder to take my medication." |
| Not seeing meds as necessary | 3 (15) | 3 | The teen does not see medicine as necessary (eg, they believe taking medicine does not help them, or they believe they do not need medicine). | "The controller one, it doesn't really do that much to you. It doesn't have an effect if you stop taking it or not. That's why it doesn't affect me that I didn't take it that much." |
| Feeling sick | 2 (10) | 3 | The teen does not take medicine due to feeling sick (eg, nauseous, having a headache). | "I could be extremely nauseous one day and not take it for the reason that I would throw up if I did." |
| Misplacement | 2 (10) | 3 | The teen does not take medicine because they or someone else lost their medicine (inhaler or spacer). | "One time [my dog] put my inhaler somewhere that I couldn't find it." |
| Feeling different | 1 (5) | 1 | The teen does not take medicine because it reminds them that their asthma diagnosis makes them different from their peers. | "It's depressing when you can't run with the other kids in gym or you can't take dance classes or you have to sit out of something or just do your own thing… You feel like you're not normal." |
| Lack of organization | 1 (5) | 1 | The teen mentions "not being organized" as a barrier to taking medicine. | "Not being organized [makes it harder to take meds], or like having so much stuff to do." |
| **External barriers** | | | | |

| Interview themes | Participants who endorsed, n (%) | Number of times a barrier was mentioned | Description of barrier | Sample quotes |
|---|---|---|---|---|
| Changes in routine | 12 (60) | 24 | The teen describes changes in their routine (eg, being away from home) as a barrier to taking medicine. | "Running late [makes it harder to take meds], and when it's not the weekdays, because in the morning I just have my routine, but on the weekends it varies." |
| Running out of meds | 11 (55) | 11 | The teen runs out of medicine because they forget to fill their prescription or they have insurance difficulties. | "I won't check how many I have left, how many pills or QVAR puffs or something, I'll run low, so I have to take one puff instead of two. Or I'll run out before we can order it and get it back in." |
| Other things get in the way | 10 (50) | 21 | The teen does not take medicine due to interference from various activities (eg, watching TV, caring for dog). These activities are often unspecified (eg, "doing other stuff"). | "Most of the time, it's hard for me to remember because I'll be thinking about other things, too, so it isn't really right there in the top of my head." |
| Friends, peers, or other people | 7 (35) | 10 | The teen does not take medicine due to influence from other people, including peers (eg, hanging out with friends, feeling victimized by peers) and siblings. | "It could be times where I hang out with friends and stuff that could get in the way [of taking meds], because I'm going to forget about everything." |
| Changes in medication | 1 (5) | 1 | The teen does not take medicine because they are in the process of changing their medicine, inhaler, or dosage. | "There was [a time] when I was in-between medicine on the one that I take every day, because they were changing the style of the inhaler and the doses, and I didn't take it for a little while." |
| Cost of medicine | 1 (5) | 1 | The teen has difficulty obtaining medicine because it is too expensive. | "[My doctor] wants to put me on a new inhaler, but the insurance doesn't want to cover it, so my pharmacy has to talk to the doctor, and they're trying to figure that out, so I'm not able to use what the doctor wants me to right now." |

### *Overview of Intervention Content Creation*

We used information gathered from individual interviews in combination with our pilot data [29] and the broader asthma literature [30,31] to develop the following ReACT intervention content: animated videos, goal-setting messages, and problem-solving modules. We created custom animated videos to provide orientation to ReACT and national guidelines–based [4] asthma education [26]. Consistent with SRT [28], we created text message libraries to prompt self-monitoring, solicit intention formation for adherence goals, and provide feedback on adherence and goal progress. Messages were gain-framed [24] and delivered based on algorithms designed to consider the recent adherence patterns of the adolescent. Finally, consistent with data gathered from individual interviews and the extant asthma management literature [31], we created problem-solving modules for 7 domains: stress, family conflict, motivation, regimen, low mood, social support, and asthma knowledge. We created text messages to guide participants to problem-solve intrapersonal and external barriers to ICS adherence endorsed via ecological momentary assessment [36]. Specifically, each module included text messages to prompt participants to identify potential goals, solutions, pros and cons, and action plans for identified barriers. Please see Figure 2 for an example problem-solving module.

**Figure 2.** Example problem-solving workflow: stress.




## Design Phase II

### *Crowdsourcing*

See Table 3 for the participant demographics for those who participated in the crowdsourcing. Notably, our crowdsourcing sample was predominantly comprised of youth from racial and ethnic minority groups (219/257, 85%). Crowdsourcing participants rated 514 out of 555 (93%) problem-solving messages as not needing modification (ie., received <40% "no" votes from participants in the crowdsourcing phase). We accepted this intervention content as semifinal, although study staff made minor revisions to this content, when appropriate, based on themes that emerged during advisory board feedback (eg, modifications to messages). Of the 555 problem-solving messages, 33 (6%) received 40.0%-59.9% "no" votes, and we revised these messages in conjunction with the advisory board. We had 8 (8/555, 1%) problem-solving messages that received >60% "no" votes. We revised both the wording and content of these messages on a case-by-case basis (see Table 4 for crowdsourcing data). Frequently, crowdsourcing participants correctly identified "problematic" solutions within the problem-solving framework and gave it a no vote (eg, "I will get angry and blame everyone else for my stress"). If participants gave suggestions to modify these messages, study staff made modifications using a consensus process.

**Table 3.** Youth, caregiver, and family demographic characteristics of the crowdsourcing participants.

| Characteristics | Youth (n=257) | Caregivers/family (n=257) |
|---|---|---|
| Age (years), mean (SD) | 15.0 (1.34) | 41.6 (7.64) |
| **Gender, n (%)** | | |
| Female | 127 (50) | 200 (78) |
| Male | 129 (50) | 57 (22) |
| Other | 1 (<1) | 0 (0) |
| **Race, n (%)** | | |
| Black/African American | 113 (44) | 113 (44) |
| Caucasian | 48 (19) | 44 (17) |
| Asian | 39 (15) | 42 (16) |
| Multiracial | 34 (13) | 36 (14)[a] |
| American Indian/Alaska Native | 18 (7) | 17 (7) |
| Native Hawaiian/Pacific Islander | 5 (2) | 5 (2) |
| **Ethnicity, n (%)** | | |
| Non-Hispanic/Latino | 207 (81) | 207 (81) |
| Hispanic/Latino | 50 (19) | 50 (19) |
| Non-Hispanic Caucasian | 38 (15) | 38 (15) |
| **Caregiver highest degree, n (%)** | | |
| Less than high school | N/A | 19 (7) |
| High school/GED[b] | N/A | 54 (21) |
| Associate degree | N/A | 55 (21) |
| Bachelor's degree | N/A | 71 (28) |
| Master's degree | N/A | 45 (17) |
| Doctorate | N/A | 5 (2) |
| Professional (eg, MD[c], JD[d]) | N/A | 4 (2) |
| Other | N/A | 4 (2) |
| **Caregiver income (US $), n (%)** | | |
| <12,000 | N/A | 20 (8) |
| 12,000-24,999 | N/A | 19 (7) |
| 25,000-49,999 | N/A | 61 (24) |
| 50,000-99,999 | N/A | 79 (31) |
| ≥100,000 | N/A | 72 28) |
| No response | N/A | 6 (2) |

[a]Caregivers were able to select more than one race; when they did, we classified them as multiracial.

[b]GED: General Educational Development.

[c]MD: Doctor of Medicine.

[d]JD: Doctor of Jurisprudence.

**Table 4.** ReACT crowdsourcing feedback about the messages.

| Votes received | Family conflict, n (%) | Knowledge, n (%) | Motivation, n (%) | Routine, n (%) | Social support, n (%) | Stress, n (%) | Low mood, n (%) | Total, n (%) |
|---|---|---|---|---|---|---|---|---|
| <40% "no" votes | 76 (95) | 68 (87) | 69 (91) | 75 (91) | 76 (93) | 69 (91) | 81 (100) | 514 (93) |
| 40%-60% "no" votes | 4 (5) | 6 (8) | 7 (9) | 7 (9) | 4 (5) | 5 (7) | 0 (0) | 33 (6) |
| >60% "no" votes | 0 (0) | 4 (5) | 0 (0) | 0 (0) | 2 (2) | 2 (3) | 0 (0) | 8 (1) |

### *Advisory Board Feedback*

During the first advisory board meeting, we introduced core ReACT functionality, provided a study timeline and overview, and described the members' role in helping to refine the intervention content and functionality of ReACT. We reviewed themes that emerged from Design Phase I and findings from the crowdsourcing data gathered as part of Design Phase II. For homework between the first 2 meetings, adolescents were asked to review and reword select intervention content that received 40%-60% "no" votes from crowdsourcing participants. During the second meeting, the group discussed the adolescents' homework responses and worked together to further refine the intervention content. Specifically, advisory board members provided suggestions on how to modify message wording (eg, using more colloquial language) and generated additional content (eg, additional solutions) for inclusion in the problem-solving modules. We dedicated the final advisory board meeting to demonstrating the features of the core ReACT intervention elements such as intention formation, feedback, problem solving, barrier identification, and motivational assessment. During this meeting, advisory board members interacted with our adherence sensor and watched the ReACT orientation and asthma education videos, and we provided them full examples via an interactive computer presentation of the ReACT interface and intervention components. The advisory board provided positive feedback regarding the look and feel of the adherence sensor. The advisory board suggested several improvements to the videos, including reducing the length of the asthma education and increasing the focus on the role of adherence to ICS in asthma management. The advisory board provided predominantly favorable feedback regarding the timing and frequency of participant interactions with ReACT. They provided suggestions on how to improve intervention messaging (eg, temper enthusiasm in messages) and sequencing.

## Discussion

### Principal Findings

Our goal with ReACT was to involve stakeholders throughout the creation and refinement of intervention content. In this way, we sought to fill a gap we identified in prior reviews of the digital intervention literature [37,38]. Consistent with best practices in mHealth intervention development [26], we used both qualitative approaches with small samples (ie., advisory boards and individual interviews) and large-sample crowdsourced feedback to develop our intervention content. The feedback we received from participants largely supported our approach. Consistent with SRT and the extant asthma literature [31,39], adolescents frequently endorsed intrapersonal and external barriers to adherence to ICS and provided generally positive impressions regarding the use of mHealth as an intervention format. We view that the variability in the types of barriers endorsed during individual interviews (eg, mood, stress, forgetfulness) as evidence that an adaptive and individually tailored mHealth system, like ReACT, may be especially beneficial for adolescents with asthma. Of note, adolescent feedback regarding barriers did not address the frequency of barriers (eg, forgetting to take medicine, feeling stressed). More research is needed to determine the barriers' frequency and impact on objective adherence behaviors.

Adolescents suggested improvements to ReACT intervention content in several instances. Similar to other intervention development studies, the main lesson learned from this process was the importance of engaging target users to increase the likelihood that our intervention content is communicated in a way that is relevant to adolescents with asthma. Refinements in intervention content came in 2 forms. First, participants helped with message clarity and ensured that our team, comprised of adult academicians, was effectively communicating behavior change concepts in language adolescents could understand [35]. Second, adolescents helped with the tone of the messages. There was a tendency among our team to generate excessively enthusiastic messages in an attempt to ensure that participants have a pleasant experience in the intervention. Our development phase revealed that some adolescents perceived this style as disingenuous. As a result, we removed exclamation points and overly enthusiastic phrases to make our messages more matter-of-fact, while still supportive.

Another important lesson learned from this development phase was that problem solving can be challenging to convey in text messages. A critical element of learning to effectively solve a problem is to briefly entertain goals and solutions that might not be productive over the long-term and may not ultimately be selected for implementation. However, our crowdsourcing participants seemed to vary in the degree to which they understood this concept. All of the messages that received a high number of "no" votes were "problematic" goals or solutions that we intended to help illustrate this component of the problem-solving process. Free-form responses made it clear that participants in the crowdsourcing were correctly identifying the message as "problematic" but appeared to misunderstand the intentional decision to include nonproductive goals and solutions as part of the problem-solving process. This confusion

may have limited the frequency and depth of feedback we received during the crowdsourcing phase.

### Limitations

We acknowledge several limitations with the current study. First, we recruited participants for individual interviews and the advisory board via convenience sampling methods. Our advisory board was comprised of a small, racially and ethnically diverse sample (n=4) of adolescents with persistent asthma from one site. Therefore, although advisory board feedback was helpful in refining messages, it is possible that participants' feedback on the intervention content is not generalizable to the larger population of adolescents with asthma. These limitations should be considered in light of involving target users in the development of ReACT from 2 study sites and gathering nationally representative feedback on ReACT intervention content from a crowdsourcing process in which 85% of youth were from racial and ethnic minority groups. Given well-documented health disparities among youth from racial/ethnic-minority youth [40], it is critical to engage a diverse sample of target end users in the design of mHealth interventions like ReACT. We posit that crowdsourcing may be a viable method to increase diversity in future mHealth intervention development studies. We acknowledge that we did not collect data on youth ICS use patterns during the crowdsourcing phase. Thus, we are unable to examine potential associations between user preferences for intervention content and self-reported ICS use. It is also noteworthy that caregiver stakeholders in the individual interview and advisory board portions were primarily female. While this likely reflects typical care patterns, we may be missing valuable perspectives from male caregivers.

### Future Research

Our immediate next steps are to conduct a pilot acceptability, usability, and preliminary efficacy study with target ReACT users [23]. Specifically, our pretest-posttest design will include a sample of 20 adolescents with persistent asthma. They will complete a 4-week baseline ICS monitoring-only period followed by a 4-week ReACT intervention period. We will gather data on enrollment rates and usage statistics and monitor technical difficulties to evaluate feasibility of ReACT. Acceptability and usability will be determined via questionnaires and a semistructured interview that asks adolescents to discuss the perceived usefulness of ReACT, how effective the intervention was in changing asthma self-management, and what changes we should make to ReACT in advance of further testing. Finally, we will evaluate preliminary efficacy by exploring changes in our hypothesized mediational variables (eg, self-efficacy) and adherence to ICS

## Acknowledgments

The National Institutes of Health supported this work (grant number 1R56HL141394-01A1).

## Authors' Contributions

DF and CC conceived the study. DF, CC, SRP, SG, EM, SP, and JS developed the protocol. DF, CC, and NK drafted the manuscript. SRP, EM, JS, and AN provided expertise and revisions. All authors completed a critical revision of the article and approved the final text.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Individual Interview Guide Questions.
[DOCX File , 18 KB - mhealth_v8i5e18400_app1.docx ]

## References

1. Centers for Disease Control and Prevention. AsthmaStats: Uncontrolled asthma among children, 2012-2014. 2012. URL: https://www.cdc.gov/asthma/asthma_stats/uncontrolled-asthma-children.htm [accessed 2020-04-25]
2. Nurmagambetov T, Kuwahara R, Garbe P. The economic burden of asthma in the United States, 2008-2013. Ann Am Thorac Soc 2018 Mar;15(3):348-356. [doi: 10.1513/AnnalsATS.201703-259OC] [Medline: 29323930]
3. Centers for Disease Control and Prevention. 2013. Asthma facts - CDC's national asthma control program grantees URL: https://www.cdc.gov/asthma/pdfs/asthma_facts_program_grantees.pdf [accessed 2020-04-26]
4. National Asthma Education and Prevention Program. Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. J Allergy Clin Immunol 2007 Nov;120(5 Suppl):S94-138. [doi: 10.1016/j.jaci.2007.09.043] [Medline: 17983880]
5. Lasmar L, Camargos P, Champs NS, Fonseca MT, Fontes MJ, Ibiapina C, et al. Adherence rate to inhaled corticosteroids and their impact on asthma control. Allergy 2009 May;64(5):784-789. [doi: 10.1111/j.1398-9995.2008.01877.x] [Medline: 19183166]

6. Williams LK, Peterson EL, Wells K, Ahmedani BK, Kumar R, Burchard EG, et al. Quantifying the proportion of severe asthma exacerbations attributable to inhaled corticosteroid nonadherence. J Allergy Clin Immunol 2011 Dec;128(6):1185-1191.e2 [FREE Full text] [doi: 10.1016/j.jaci.2011.09.011] [Medline: 22019090]
7. McQuaid EL, Penza-Clyve SM, Nassau JH, Fritz GK, Klein R, O'Connor S, et al. The Asthma Responsibility Questionnaire: Patterns of Family Responsibility for Asthma Management. Children's Health Care 2001 Sep;30(3):183-199. [doi: 10.1207/S15326888CHC3003_2]
8. McQuaid EL, Kopel SJ, Klein RB, Fritz GK. Medication adherence in pediatric asthma: reasoning, responsibility, and behavior. J Pediatr Psychol 2003;28(5):323-333. [doi: 10.1093/jpepsy/jsg022] [Medline: 12808009]
9. Naimi DR, Freedman TG, Ginsburg KR, Bogen D, Rand CS, Apter AJ. Adolescents and asthma: why bother with our meds? J Allergy Clin Immunol 2009 Jun;123(6):1335-1341 [FREE Full text] [doi: 10.1016/j.jaci.2009.02.022] [Medline: 19395075]
10. Drotar D, Bonner MS. Influences on adherence to pediatric asthma treatment: a review of correlates and predictors. J Dev Behav Pediatr 2009 Dec;30(6):574-582. [doi: 10.1097/DBP.0b013e3181c3c3bb] [Medline: 19996903]
11. D'Zurilla TJ, Maydeu-Olivares A, Kant GL. Age and gender differences in social problem-solving ability. Personality and Individual Differences 1998 Aug;25(2):241-252. [doi: 10.1016/s0191-8869(98)00029-4]
12. Jaffee WB, D'Zurilla TJ. Adolescent problem solving, parent problem solving, and externalizing behavior in adolescents. Behavior Therapy 2003;34(3):295-311. [doi: 10.1016/s0005-7894(03)80002-3]
13. Muscara F, Catroppa C, Anderson V. Social problem-solving skills as a mediator between executive function and long-term social outcome following paediatric traumatic brain injury. J Neuropsychol 2008 Sep;2(2):445-461. [doi: 10.1348/174866407x250820] [Medline: 19824165]
14. McGee CL, Fryer SL, Bjorkquist OA, Mattson SN, Riley EP. Deficits in social problem solving in adolescents with prenatal exposure to alcohol. Am J Drug Alcohol Abuse 2008 Jul 07;34(4):423-431. [doi: 10.1080/00952990802122630] [Medline: 18584572]
15. Lansing AH, Berg CA. Adolescent self-regulation as a foundation for chronic illness self-management. J Pediatr Psychol 2014;39(10):1091-1096 [FREE Full text] [doi: 10.1093/jpepsy/jsu067] [Medline: 25214646]
16. Wolf FM, Guevara JP, Grum CM, Clark NM, Cates CJ. Educational interventions for asthma in children. Cochrane Database Syst Rev 2003;4(1):CD000326. [doi: 10.1002/14651858.CD000326] [Medline: 12535395]
17. Bruzzese J, Bonner S, Vincent EJ, Sheares BJ, Mellins RB, Levison MJ, et al. Asthma education: the adolescent experience. Patient Educ Couns 2004 Dec;55(3):396-406. [doi: 10.1016/j.pec.2003.04.009] [Medline: 15582346]
18. Jonsson M, Bergström A, Egmar A, Hedlin G, Lind T, Kull I. Asthma during adolescence impairs health-related quality of life. J Allergy Clin Immunol Pract 2016 Jan;4(1):144-6.e2. [doi: 10.1016/j.jaip.2015.07.020] [Medline: 26342743]
19. Cui W, Zack MM, Zahran HS. Health-related quality of life and asthma among United States adolescents. J Pediatr 2015 Feb;166(2):358-364 [FREE Full text] [doi: 10.1016/j.jpeds.2014.10.005] [Medline: 25454936]
20. Ramsey RR, Plevinsky JM, Kollin SR, Gibler RC, Guilbert TW, Hommel KA. Systematic Review of Digital Interventions for Pediatric Asthma Management. J Allergy Clin Immunol Pract 2020 Apr;8(4):1284-1293. [doi: 10.1016/j.jaip.2019.12.013] [Medline: 31870809]
21. Bender BG. Nonadherence to Asthma Treatment: Getting Unstuck. J Allergy Clin Immunol Pract 2016;4(5):849-851. [doi: 10.1016/j.jaip.2016.07.007] [Medline: 27587318]
22. Ramsey RR, Caromody JK, Voorhees SE, Warning A, Cushing CC, Guilbert TW, et al. A Systematic Evaluation of Asthma Management Apps Examining Behavior Change Techniques. J Allergy Clin Immunol Pract 2019;7(8):2583-2591. [doi: 10.1016/j.jaip.2019.03.041] [Medline: 30954644]
23. Cushing CC, Fedele DA, Patton SR, McQuaid EL, Smyth JM, Prabhakaran S, et al. Responsive Asthma Care for Teens (ReACT): development protocol for an adaptive mobile health intervention for adolescents with asthma. BMJ Open 2019 Aug 20;9(8):e030029 [FREE Full text] [doi: 10.1136/bmjopen-2019-030029] [Medline: 31434777]
24. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
25. 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]
26. Fedele D, McConville A, Moon J, Thomas J. Topical Review: Design Considerations When Creating Pediatric Mobile Health Interventions: Applying the IDEAS Framework. J Pediatr Psychol 2019 Apr 01;44(3):343-348 [FREE Full text] [doi: 10.1093/jpepsy/jsy084] [Medline: 30346561]
27. Centers for Disease Control and Prevention. Asthma - National Health Interview Survey (NHIS) data URL: https://www.cdc.gov/asthma/nhis/default.htm [accessed 2020-04-27]
28. Clark NM, Gong M, Kaciroti N. A model of self-regulation for control of chronic disease. Health Educ Behav 2001 Dec 30;28(6):769-782. [doi: 10.1177/109019810102800608] [Medline: 11720277]
29. Netz M, Fedele DA, Sweenie R, Baker D, Light M, McQuaid EL. The role of youth responsibility for disease management in quality of life among early adolescents with asthma. J Pediatr Psychol 2019;45(1):40-49.

30. Gray WN, Netz M, McConville A, Fedele D, Wagoner ST, Schaefer MR. Medication adherence in pediatric asthma: A systematic review of the literature. Pediatr Pulmonol 2018 May 20;53(5):668-684. [doi: 10.1002/ppul.23966] [Medline: 29461017]
31. Mammen J, Rhee H. Adolescent Asthma Self-Management: A Concept Analysis and Operational Definition. Pediatr Allergy Immunol Pulmonol 2012 Dec;25(4):180-189 [FREE Full text] [doi: 10.1089/ped.2012.0150] [Medline: 23285426]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Corbin J, Strauss A. Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks, California: SAGE Publications; 2008.
34. Miles M, Huberman A, Sandaña J. Qualitative data analysis: A methods sourcebook. 3rd ed. Thousand Oaks, CA: Sage; 2013.
35. Woolford SJ, Barr KLC, Derry HA, Jepson CM, Clark SJ, Strecher VJ, et al. OMG do not say LOL: obese adolescents' perspectives on the content of text messages to enhance weight loss efforts. Obesity (Silver Spring) 2011 Dec;19(12):2382-2387 [FREE Full text] [doi: 10.1038/oby.2011.266] [Medline: 21869762]
36. D'Zurilla TJ, Nezu AM. Problem-Solving Therapy. In: Dobson KS, editor. Handbook of Cognitive-Behavioral Therapies. 3rd ed. New York, NY: The Guilford Press; 2010:197-225.
37. Fedele DA, Cushing CC, Fritz A, Amaro CM, Ortega A. Mobile Health Interventions for Improving Health Outcomes in Youth: A Meta-analysis. JAMA Pediatr 2017 May 01;171(5):461-469 [FREE Full text] [doi: 10.1001/jamapediatrics.2017.0042] [Medline: 28319239]
38. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May 13;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
39. Clark NM, Shah S, Dodge JA, Thomas LJ, Andridge RR, Little RJA. An evaluation of asthma interventions for preteen students. J Sch Health 2010 Feb;80(2):80-87 [FREE Full text] [doi: 10.1111/j.1746-1561.2009.00469.x] [Medline: 20236406]
40. Canino G, McQuaid EL, Rand CS. Addressing asthma health disparities: a multilevel challenge. J Allergy Clin Immunol 2009 Jun;123(6):1209-17; quiz 1218 [FREE Full text] [doi: 10.1016/j.jaci.2009.02.043] [Medline: 19447484]


## Abbreviations

**GED:** General Educational Development.
**ICS:** inhaled corticosteroids.
**JD:** Doctor of Jurisprudence.
**MD:** Doctor of Medicine.
**mHealth:** mobile health.
**ReACT:** Responsive Asthma Care for Teens.
**SRT:** self-regulation theory.


*Edited by G Eysenbach; submitted 24.02.20; peer-reviewed by C Stiles-Shields; comments to author 18.03.20; revised version received 27.03.20; accepted 08.04.20; published 06.05.20.*

*Please cite as:*
*Fedele DA, Cushing CC, Koskela-Staples N, Patton SR, McQuaid EL, Smyth JM, Prabhakaran S, Gierer S, Nezu AM*
*Adaptive Mobile Health Intervention for Adolescents with Asthma: Iterative User-Centered Development*
*JMIR Mhealth Uhealth 2020;8(5):e18400*
*URL: https://mhealth.jmir.org/2020/5/e18400*
*doi:10.2196/18400*
*PMID:32374273*

Original Paper

# Efficacy and Impact of Digital HIV Care Navigation in Young People Living With HIV in San Francisco, California: Prospective Study

Sean Arayasirikul[1,2,3], PhD; Caitlin Turner[1], MPH; Dillon Trujillo[1], MPH; Victory Le[1], MA; Erin C Wilson[1], DrPH

[1]San Francisco Department of Public Health, Center for Public Health Research, San Francisco, CA, United States

[2]Department of Pediatrics, University of California San Francisco, San Francisco, CA, United States

[3]Department of Psychiatry, University of California San Francisco, San Francisco, CA, United States

**Corresponding Author:**
Sean Arayasirikul, PhD
San Francisco Department of Public Health
Center for Public Health Research
25 Van Ness Avenue
San Francisco, CA,
United States
Phone: 1 415 554 9000
Email: sean.arayasirikul@gmail.com

## Abstract

**Background:** Young people are disproportionately impacted by HIV infection and exhibit poor HIV care continuum outcomes. Mobile health (mHealth) interventions are promising approaches to meet the unique needs of young people living with HIV. Youth-focused interventions are needed to improve HIV care continuum outcomes.

**Objective:** This study assessed the preliminary efficacy and impact of a digital HIV care navigation intervention among young people living with HIV in San Francisco. Health electronic navigation (eNavigation or eNav) is a 6-month, text message–based, digital HIV care navigation intervention, in which young people living with HIV are connected to their own HIV care navigator through text messaging to improve engagement in HIV primary care.

**Methods:** This study had a single-arm, prospective, pre-post design. The analysis included 120 young men who have sex with men or transwomen living with HIV aged between 18 and 34 years. We analyzed self-reported sociobehavioral information pre- and postintervention at baseline and 6 months, which was collected using computer-assisted self-interviewing surveys. We characterized the sample and built generalized estimating equation (GEE) models to assess differences in HIV care continuum outcomes at baseline and 6 months.

**Results:** The characteristics according to the intervention completion status were not different from those of the overall sample. The mean age of the participants was 27.75 years (SD 4.07). Most participants (103/120, 85.8%) identified as men, and the sample was racially/ethnically diverse. At baseline, majority (99/120, 82.5%) of the participants had recently received primary HIV care, yet this was more likely in those who completed the intervention than in those who did not (54/60, 90% vs 45/60, 75%; $\chi^2_1$=4.68, *P*=.03). More than half of the sample reported taking antiretroviral therapy (92/120, 76.7%) and having an undetectable viral load (65/120, 54.2%). The 6-month follow-up surveys were completed by 73.3% (88/120) of participants, and these participants were not characteristically different from the overall sample at baseline. GEE models indicated that participants had increased odds of viral suppression at 6 months as compared with baseline. No relevant additive or multiplicative interactions were noted on comparing outcome effects over time according to intervention completion.

**Conclusions:** Digital HIV care navigation fills a critical gap in public health and HIV care systems, making these systems more responsive and accountable to the needs of the most vulnerable individuals. Our intervention bridges the time between primary care visits with interactive, tailored, personalized, and peer-delivered social support; information; and motivational interviewing to scaffold behavioral change. This study is part of the next wave of system-informed mHealth intervention research that will offer potentially disruptive solutions to traditional in-person delivered interventions and improve the health of the most vulnerable individuals.

**International Registered Report Identifier (IRRID):** RR2-10.2196/16406

*(JMIR Mhealth Uhealth 2020;8(5):e18597)* doi:10.2196/18597

**KEYWORDS**

HIV/AIDS; digital HIV care navigation; young people living with HIV; mHealth

## Introduction

### Background

HIV remains a pressing public health issue, with over a million people in the United States living with HIV [1]. Although effective antiretroviral therapy (ART) is available to prevent HIV transmission from people living with HIV, disparities persist, and young people are affected. Even though the number of new infections among people aged 13 to 24 years decreased between 2010 and 2016, HIV transmission among young adults aged 25 to 34 years has increased [2]. Youth and young adults are not only disproportionately at risk for HIV but also experience relevant gaps in the HIV care continuum. Data from the Ryan White HIV/AIDS Program show that the viral suppression rate is less than 70% and is highest among those aged 25 to 30 years (68%), followed by those aged 13 to 18 years (66%) and aged 19 to 24 years (59%) [3]. Youth-focused interventions are needed to improve HIV care continuum outcomes.

Mobile health (mHealth) interventions have greatly increased over the last decade. A systematic review of systematic reviews identified 23 systematic reviews representing more than 10,000 scientific articles and involving close to 80,000 participants in 371 mHealth studies [4]. Yet, evidence for the efficacy and impact of mHealth interventions is unclear. This study assessed the preliminary efficacy and impact of a digital HIV care navigation intervention among young people living with HIV in San Francisco.

### Overview of Health Electronic Navigation

Health electronic navigation (eNavigation or eNav) is a 6-month, text message–based, digital HIV care navigation intervention, in which young people living with HIV are connected to their own HIV care navigator through text messaging to improve engagement in HIV primary care. The intervention includes delivery of the following: (1) HIV care navigation, (2) health promotion and education, (3) motivational interviewing, and (4) social support. HIV care navigation guides participants in knowing where, when, and how to access all health and related services, and increases access to appropriate resources (eg, primary medical care, mental health care, housing, insurance, and benefits) [5]. Health promotion and education ensure optimal health literacy for all participants by providing information on HIV biology, disease management, communication with providers, risk reduction, healthy behavior, and ART adherence. Motivational interviewing is a technique and a style of counseling that can help resolve the ambivalence that prevents patients from realizing their personal goals [6,7]. Social support is provided through the establishment of an open nonjudgmental care relationship between participants and their HIV care navigator to address life events and topics most important to young people living with HIV that may not be solely focused on their HIV care.

## Methods

### Ethics Approval

All procedures were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study protocol was approved by the Institutional Review Board of the University of California, San Francisco (IRB #16-19675).

### Study Design and Recruitment

Data for the present analysis were collected from the Health eNav study at San Francisco Department of Public Health (2017-2018). Health eNav is a digital care navigation intervention designed to improve HIV care continuum outcomes among young men who have sex with men and transwomen living with HIV. A digital care navigator delivered the intervention via two-way SMS text messaging. This study had a single-arm, prospective, pre-post design. Study procedures are described in-depth in a prior manuscript [8].

Eligibility criteria for the study were as follows: identifying as a man who has sex with men or a transwoman; being between the ages of 18 and 34 years; and being newly diagnosed with HIV, not being engaged/retained in HIV care, or having a detectable viral load. Participants were recruited via convenience sampling from five clinics and community-based organizations in San Francisco serving young people living with HIV. If eligible, participants visited research staff at study offices within the San Francisco Department of Public Health, where informed consent was obtained. Of 170 individuals screened, 140 were eligible. However, 20 were subsequently lost to follow-up and were not enrolled. The final sample had 120 young men who have sex with men or transwomen living with HIV.

### Data Collection and Measures

Data for this analysis were collected using computer-assisted self-interviewing (CASI) surveys. Instruments collected self-reported sociobehavioral information pre- and postintervention at baseline and 6 months.

#### *Sociodemographic Information*

We assessed age at interview (in years), gender identity (transwoman vs man), race/ethnicity (non-Hispanic/Latinx, American Indian/Alaska Native, Asian, black/African American, multiracial, white, or Hispanic/Latinx), and education level (high school/General Educational Development [GED] or at least college education). We also examined the current living situation, defined as stable (owning or renting a house or apartment) or unstable (living with someone who owns or rents a home, living in temporary or transitional housing, or experiencing homelessness). Income was classified as earning US $0 to 250, US $251 to 600, US $601 to 1300, or US $1301

or more in the last month. Finally, we assessed history of incarceration.

### *Intervention Completion Status*

We defined intervention completion as retention in the 6-month intervention. Among the 120 participants enrolled in the intervention, 60 were lost to follow-up and did not complete the 6-month intervention. The most common reason for loss to follow-up during the intervention was unknown (40/60, 67%), followed by participant phone loss (6/60, 10%), moving out of jurisdiction (9/60, 15%), incarceration (2/60, 3%), withdrawal (2/60, 3%), and death from causes unrelated to intervention participation (1/60, 2%). For participants who lost their phones, a maximum of one phone replacement was provided during the intervention. Participants who again lost their phones were unable to complete the intervention. The intervention completion status was dichotomously coded as "yes (1)" if participants completed the 6-month intervention and "no (0)" if they did not complete the intervention.

### *HIV Care Continuum Outcomes*

Self-reported HIV care outcome data were collected using CASI surveys. In accordance with HIV care goals designated in 90-90-90 objectives [9], we dichotomized outcomes regarding whether participants received primary HIV care within the last 6 months (yes/no, 1/0), whether participants were taking ART (yes/no, 1/0), and whether participants were virally suppressed (eg, had a viral load of 200 copies/mL or less) (yes/no, 1/0).

### Statistical Analysis

Initially, we characterized the sample by describing baseline sociodemographics and self-reported HIV care continuum outcomes using baseline CASI data. Differences in baseline sociodemographics and outcomes on comparing participants who completed the intervention and those who did not were analyzed with bivariate statistical tests (chi-squared test or *t* test). As every participant received the digital care navigation intervention, we analyzed HIV care continuum outcome effects for all participants over the 6-month follow-up period. We produced three models using generalized estimating equations (GEEs) to account for within-subject correlations and to calculate the odds of HIV care continuum outcomes for the 6-month follow-up compared with baseline.

Finally, we assessed whether intervention completion modified HIV care continuum outcomes over time. To accomplish this, we tested for multiplicative and additive interactions between intervention completion and time (baseline or 6-month interview) and computed stratum-specific results. Initially, outcome odds or prevalences from the GEE models with interaction terms were converted to probabilities. We then calculated probability ratios for multiplicative interactions (ie, probability of having an undetectable viral load over the 6-month period for those who completed the intervention compared with those who did not complete the intervention), probability differences for additive interactions (ie, excess probability of having an undetectable viral load over the 6-month time period for those who completed the intervention compared with those who did not complete the intervention), and stratum-specific differences (ie, excess probability of having an undetectable viral load over the 6-month time period only among those who completed the intervention or only among those who did not complete the intervention). We reported statistically significant stratum-specific results and statistically significant interactions between intervention completion and time. All statistical analyses were performed using Stata 14 (StataCorp LLC College Station, Texas, USA) [10]; comparisons were considered statistically significant if the associated *P* value was less than .05.

## Results

Table 1 presents baseline sociodemographics and HIV care continuum outcomes for the Health eNav sample overall (n=120) and according to the intervention completion status. Except for incarceration and recent receipt of primary HIV care, characteristics according to the intervention completion status were not significantly different from those of the overall sample. The mean age of the participants was 27.75 years (SD 4.07). Most participants (103/120, 85.8%) identified as men. The sample was racially/ethnically diverse, with most participants identifying as Hispanic/Latinx, followed by white, multiple races, and black/African American, and few identifying as Asian or American Indian/Alaska Native. About half (68/120, 56.7%) of all participants completed some college education, yet most lived in unstable housing and had a monthly income of US $1300 or less. Recent incarceration was less likely in participants who completed the intervention than those who did not complete the intervention (11.67% vs. 26.67%, $\chi^2_1$=4.36, *P*=.04).

In terms of baseline HIV care continuum outcomes, majority (99/120, 82.5%) of the participants had recently received primary HIV care, yet this was more likely in those who completed the intervention than in those who did not complete the intervention (54/60, 90% vs 45/60, 75%; $\chi^2_1$=4.68, *P*=.03). More than half of the sample reported taking ART (92/120, 76.7%) and having an undetectable viral load (65/120, 54.2%) (Table 1).

The 6-month follow-up surveys were completed by 73.3% (88/120) of participants (Table 2), and these participants were not characteristically different from the overall sample at baseline. Table 2 presents the longitudinal results from the GEE models. After analyzing HIV care continuum outcomes over the 6-month study period, we observed that participants had increased odds of viral suppression at 6 months compared with baseline. We observed no statistically significant additive or multiplicative interactions on comparing outcome effects over time according to intervention completion. However, on testing for stratum-specific effects, we found that viral suppression increased over time among those who completed the intervention (83.89% probability of viral suppression at 6 months vs 69.60% probability of viral suppression at baseline; probability difference 14.29%, 95% CI 2.66%-26.41%). No corresponding stratum-specific difference in viral suppression was observed among those who did not complete the intervention.

**Table 1.** Differences in baseline sociodemographics and HIV care continuum outcomes among young men who have sex with men and transwomen overall and according to intervention completion (Health eNav, 2017-2019).

| Characteristic | Overall, n (%) or mean (SD) | Did not complete the intervention, n (%) or mean (SD) | Completed the intervention, n (%) or mean (SD) | Group comparison | |
|---|---|---|---|---|---|
| | | | | *t* test or chi-square test (df=1) statistic | *P* value |
| Total | 120 (100.0) | 60 (50.0) | 60 (50.0) | —[a] | — |
| **Sociodemographics** | | | | | |
| Age (years) | 27.75 (4.07) | 27.57 (4.09) | 27.93 (4.07) | *t*=0.49 | .62 |
| **Gender identity** | | | | **$\chi^2$=0.07** | **.79** |
| Transwoman | 17 (14.2) | 9 (15.0) | 8 (13.3) | | |
| Man | 103 (85.8) | 51 (85.0) | 52 (86.7) | | |
| **Race/ethnicity** | | | | **$\chi^2$=4.42** | **.22** |
| Black, non-Hispanic/Latinx | 22 (18.3) | 11 (18.3) | 11 (18.3) | | |
| Hispanic/Latinx | 38 (31.7) | 24 (40.0) | 14 (23.3) | | |
| Multiple races, non-Hispanic/Latinx | 28 (23.3) | 11 (18.3) | 17 (28.3) | | |
| White, non-Hispanic/Latinx | 32 (26.7) | 14 (23.3) | 18 (30.0) | | |
| **Education level** | | | | **$\chi^2$=3.39** | **.07** |
| High school/GED[b] or less | 52 (43.3) | 31 (51.7) | 21 (35.0) | | |
| Some college or more | 68 (56.7) | 29 (48.3) | 39 (65.0) | | |
| **Current living situation** | | | | **$\chi^2$=3.08** | **.08** |
| Unstable | 81 (67.5) | 45 (75.0) | 36 (60.0) | | |
| Stable | 39 (32.5) | 15 (25.0) | 24 (40.0) | | |
| **Income in the last month (US $)** | | | | **$\chi^2$=0.16** | **.98** |
| 601-1300 | 30 (25.0) | 15 (25.0) | 15 (25.0) | | |
| 251-600 | 30 (25.0) | 14 (23.3) | 16 (26.7) | | |
| 0-250 | 30 (25.0) | 15 (25.0) | 15 (25.0) | | |
| 1301 or more | 29 (24.2) | 15 (25.0) | 14 (23.3) | | |
| **Incarceration** | | | | **$\chi^2$=4.36** | **.04** |
| Yes | 23 (19.2) | 16 (26.7) | 7 (11.7) | | |
| No | 97 (80.8) | 44 (73.3) | 53 (88.3) | | |
| **HIV care continuum outcomes** | | | | | |
| **Received primary HIV care, last 6 months** | | | | **$\chi^2$=4.68** | **.03** |
| Yes | 99 (82.5) | 45 (75.0) | 54 (90.0) | | |
| No | 21 (17.5) | 15 (25.0) | 6 (10.0) | | |
| **Currently taking ART[c]** | | | | **$\chi^2$=3.69** | **.06** |
| Yes | 92 (76.7) | 42 (70.0) | 50 (83.3) | | |
| No | 27 (22.5) | 18 (30.0) | 9 (15.0) | | |
| **Undetectable viral load** | | | | **$\chi^2$=0.13** | **.72** |
| Yes | 65 (54.2) | 28 (46.7) | 37 (61.7) | | |
| No | 32 (26.7) | 15 (25.0) | 17 (28.3) | | |

[a]Not applicable.

[b]GED: General Educational Development.

[c]ART: antiretroviral therapy.

**Table 2.** Differences in HIV care continuum outcomes at baseline and 6 months among men who have sex with men and transwomen living with HIV (Health eNav, 2017-2019).

| Characteristic | Baseline, n (%) | 6 months, n (%) | Percentage mean change | Generalized estimating equation[a] outcome effects over time | |
|---|---|---|---|---|---|
| | | | | OR (95% CI) | *P* value |
| Total | 120 (100.0) | 88 (73.3) | —[b] | — | — |
| **HIV care continuum outcomes** | | | | | |
| **Received primary HIV care, last 6 months** | | | | | |
| No | 21 (17.5) | 9 (10.2) | — | Reference | — |
| Yes | 99 (82.5) | 79 (89.8) | 7.3 | 1.79 (0.82-3.93) | .14 |
| **Currently taking ART[c]** | | | | | |
| No | 27 (22.5) | 16 (18.2) | — | Reference | — |
| Yes | 92 (76.7) | 72 (81.8) | 5.2 | 1.13 (0.70-1.83) | .61 |
| **Undetectable viral load** | | | | | |
| No | 32 (26.7) | 13 (14.8) | — | Reference | — |
| Yes | 65 (54.2) | 67 (76.1) | 22.0 | 2.16 (1.30-3.57) | .003 |

[a]Three generalized estimating equation models were used for the analyses: (1) odds of receiving primary HIV care at 6 months compared with baseline, (2) odds of taking ART at 6 months compared with baseline, and (3) odds of being virally suppressed at 6 months compared with baseline.

[b]Not applicable.

[c]ART: antiretroviral therapy.

# Discussion

## Principal Findings

Our findings suggest that digital HIV care navigation is effective at promoting viral suppression at post-test compared with pretest. While many mHealth interventions developed for youth and young adults living with HIV are increasingly utilizing methods that leverage automated functionalities (eg, reminders, calendaring, gamification, and diary studies), Health eNav demonstrates the importance of real-time bidirectional interaction with an interventionist using text messaging. In other research, we found that digital HIV care navigation is feasible and acceptable and, in particular, responsive to the diverse needs of young people living with HIV in a metropolitan city with complex structural barriers [11]. Despite the lack of precedent, we argue that mHealth interventions must be scalable beyond the individual level in order to strengthen health systems and public health action [12]. While local health departments may be weary of integrating mHealth interventions into their system-wide approaches, this study provides evidence for the positive impact that such work can have on individuals in the health system [13].

## Limitations and Future Research

The findings should be interpreted with some limitations in mind. This study had a single-arm, prospective, pre-post design. Future studies using more robust study designs with randomization of participants to multiple study arms are needed to develop a cogent understanding of the short- and long-term efficacies of digital HIV care navigation. Additionally, future analyses that examine dose-response or employ just-in-time stepped wedge designs may offer more dynamic and responsive flexibility to the study population of young men who have sex with men and transwomen living with HIV. These individuals not only are the most vulnerable to HIV acquisition and the most likely to have poor HIV care continuum outcomes, but also face multiple structural barriers and complex stigma. While adolescence and young adulthood are already dynamic critical periods, for sexual and gender minorities living with HIV, the situation is exacerbated by intersectional stigma. More research on the understanding of how the critical axes of race, sexuality, gender identity, and HIV interact is needed. mHealth interventions addressing intersectional stigma must be developed.

## Conclusion

Despite the study limitations, digital HIV care navigation fills a critical gap in public health and HIV care systems, making these systems more responsive and accountable to the needs of the most vulnerable individuals. Our intervention bridges the time between primary care visits with interactive, tailored, personalized, peer-delivered social support; information; and motivational interviewing to scaffold behavioral changes, such as ART adherence. This study is part of the next wave of system-informed mHealth intervention research that will offer potentially disruptive solutions to traditional in-person delivered interventions and improve the health of the most vulnerable individuals. For ending the HIV epidemic in San Francisco, the United States, and globally, novel applications of low-tech hi-touch mHealth technologies, such as digital HIV care navigation, should be considered.

## Conflicts of Interest

None declared.

## References


1. Hess K, Johnson S, Hu X, Li J, Wu B, Yu C. Centers for Disease Control and Prevention HIV Surveillance Report, 2017. 2018 Nov. Diagnoses of HIV infection in the United States and dependent areas, 2017 URL: https://stacks.cdc.gov/view/cdc/60911 [accessed 2020-04-15]
2. HIV.gov. Atlanta, Georgia: Centers for Disease Control and Prevention; 2020 Jan 16. HIV Overview: Data & Trends, U.S. Statistics URL: https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics [accessed 2020-02-01]
3. Health Resources and Services Administration (HRSA). Youth and Young Adults and the Ryan White HIV/AIDS Program. Health Resources and Services Administration (HRSA) 2015.
4. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
5. Higa DH, Marks G, Crepaz N, Liau A, Lyles CM. Interventions to improve retention in HIV primary care: a systematic review of U.S. studies. Curr HIV/AIDS Rep 2012 Dec;9(4):313-325 [FREE Full text] [doi: 10.1007/s11904-012-0136-6] [Medline: 22996171]
6. Davidson R. Motivational issues in the treatment of addictive behaviour. Psychotherapy, Psychological Treatments and the Addictions 1996 Mar 07:173-188. [doi: 10.1017/CBO9780511722127.012]
7. Miller W, Rollnick S. Motivational Interviewing: Helping People Change, 3rd Edition (Applications Of Motivational Interviewing). New York City, New York: The Guilford Press; 2020.
8. Arayasirikul S, Trujillo D, Turner CM, Le V, Wilson EC. Implementing a Digital HIV Care Navigation Intervention (Health eNav): Protocol for a Feasibility Study. JMIR Res Protoc 2019 Nov 08;8(11):e16406 [FREE Full text] [doi: 10.2196/16406] [Medline: 31702561]
9. Del Rio C. Editorial: Can we end HIV as a public health problem globally? Progress towards achieving the UNAIDS 90-90-90 goals. Curr Opin HIV AIDS 2019 Nov;14(6):439-441. [doi: 10.1097/COH.0000000000000592] [Medline: 31580265]
10. StataCorp. Stata Statistical Software: Release 14. College Station, Texas: StataCorp LP; 2015.
11. Trujillo D, Turner C, Le V, Wilson E, Arayasirikul S. Digital HIV Care Navigation for Young People Living With HIV in San Francisco, California: Feasibility and Acceptability Study. JMIR Mhealth Uhealth 2020 Jan 10;8(1):e16838 [FREE Full text] [doi: 10.2196/16838] [Medline: 31922489]
12. Malvey D, Slovensky D. Putting mHealth in Public Health. In: mHealth. Boston, Massachusetts: Springer; 2014:115-139.
13. Rubin S, Bouri N, Jolani N, Minton K. The adoption of social media and mobile health technologies for emergency preparedness by local health departments: a joint perspective from NACCHO and the UPMC center for health security. J Public Health Manag Pract 2014;20(2):259-263. [doi: 10.1097/PHH.0000000000000056] [Medline: 24458314]


## Abbreviations

**ART:** antiretroviral therapy
**CASI:** computer-assisted self-interviewing
**GEE:** generalized estimating equation
**mHealth:** mobile health

*Edited by G Eysenbach; submitted 06.03.20; peer-reviewed by M Thompson; comments to author 27.03.20; revised version received 31.03.20; accepted 31.03.20; published 08.05.20.*

*Please cite as:*
*Arayasirikul S, Turner C, Trujillo D, Le V, Wilson EC*
*Efficacy and Impact of Digital HIV Care Navigation in Young People Living With HIV in San Francisco, California: Prospective Study*
*JMIR Mhealth Uhealth 2020;8(5):e18597*
*URL: https://mhealth.jmir.org/2020/5/e18597*
*doi:10.2196/18597*
*PMID:32383680*

Original Paper

# Development of the InCharge Health Mobile App to Improve Adherence to Hydroxyurea in Patients With Sickle Cell Disease: User-Centered Design Approach

Nicole M Alberts[1*], PhD; Sherif M Badawy[2,3*], MS, MD, MBBCh; Jason Hodges[4], MA, PhD; Jeremie H Estepp[4], MD; Chinonyelum Nwosu[4,5], BS, MPH; Hamda Khan[4,5], BS, MA; Matthew P Smeltzer[6], PhD; Ramin Homayouni[7], BS, PhD; Sarah Norell[8], MFA, MDes; Lisa Klesges[6], BA, MS, PhD; Jerlym S Porter[1*], MPH, PhD; Jane S Hankins[4,5*], MS, MD

[1]Department of Psychology, St Jude Children's Research Hospital, Memphis, TN, United States

[2]Department of Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[3]Division of Hematology, Oncology and Stem Cell Transplant, Ann & Robert H Lurie Children's Hospital of Chicago, Chicago, IL, United States

[4]Department of Hematology, St Jude Children's Research Hospital, Memphis, TN, United States

[5]Methodist Health Care, Memphis, TN, United States

[6]Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN, United States

[7]Department of Foundational Medical Studies, William Beaumont School of Medicine, Oakland University, Rochester, MI, United States

[8]University of Illinois Health, Institute for Healthcare Delivery Design, Chicago, IL, United States

[*]these authors contributed equally

**Corresponding Author:**
Jane S Hankins, MS, MD
Department of Hematology
St Jude Children's Research Hospital
332 Danny Thomas Place, Mail Stop 800
Memphis, TN, 38105
United States
Phone: 1 901 595 4153
Email: Jane.Hankins@stjude.org

## Abstract

**Background:** Sickle cell disease (SCD) is an inherited blood disorder causing acute complications and chronic progressive end organ damage. SCD is associated with significant morbidity, early mortality, impaired health-related quality of life, and increased acute health care utilization. Hydroxyurea is a US Food and Drug Administration–approved medication that reduces disease complications, acute health care utilization, and costs. However, adherence to hydroxyurea is suboptimal. Mobile health (mHealth) interventions have the potential to improve hydroxyurea adherence, but few examples exist that are specific to the SCD population.

**Objective:** This study aimed to design a mHealth intervention for individuals with SCD to improve adherence to hydroxyurea, using a user-centered design that was informed by specific barriers to hydroxyurea adherence and utilization in this population.

**Methods:** This study consisted of 4 phases. In phase 1, individuals with SCD and health care providers participated in an optimization digital workshop. In phase 2, patients completed surveys pertaining to their interest in mHealth use, barriers and facilitators to hydroxyurea use, and health literacy. Phases 3 and 4 involved semistructured interviews and focus groups, respectively, and used the Health Belief Model (HBM) as the framework to investigate drivers of poor hydroxyurea adherence and to inform the development of an app prototype. In addition, in phase 4, we have incorporated the patients' feedback on the preliminary app prototype and its features.

**Results:** Barriers to hydroxyurea adherence were consistent with the literature and included forgetfulness and several specific thoughts and emotions associated with hydroxyurea use (eg, fear of side effects, depression, stigma, and hopelessness). In addition, more than half of the participants reported potentially low health literacy. Preferred patient app features included 7 key components, namely (1) medication reminders and tracker, (2) disease education, (3) communication, (4) personalization, (5) motivation, (6) support during pain episodes, and (7) social support. Utilizing a user-centered design approach, data obtained from patients and providers were translated into features within the app, mapping to components of the HBM and the specific drivers of hydroxyurea

adherence and matching the literacy level of the population, resulting in the development of a novel mobile app called InCharge Health.

**Conclusions:** The InCharge Health app is an mHealth intervention developed with substantial input from users and by mapping the HBM as the framework that guided the choice for its components. InCharge Health is a customized product for the SCD population aimed at optimizing medication adherence, with the end goal of improving quality of life and health outcomes among patients with SCD. The efficacy and implementation of the InCharge Health app as an mHealth intervention to promote hydroxyurea adherence will be tested in a future stepped-wedge multicenter trial for adolescents and adults with SCD.

***(JMIR Mhealth Uhealth 2020;8(5):e14884)*** doi:10.2196/14884

**KEYWORDS**

sickle cell anemia; hydroxyurea; hydroxycarbamide; medication adherence; self-management; mobile health; internet

## Introduction

### Background

Sickle cell disease (SCD) is a common inherited hemoglobin disorder in the United States, affecting approximately 100,000 Americans, mainly of African American descent [1-3]. SCD is a debilitating illness with several acute and chronic complications, including vaso-occlusive pain episodes, acute chest syndrome, cerebrovascular events, cognitive dysfunction, and progressive end organ damage. SCD has been associated with significant morbidity, early mortality, impaired health-related quality of life [3-5], and increased health care utilization [2,6]. Hydroxyurea (also known as hydroxycarbamide) is a medication approved by the US Food and Drug Administration (FDA) that has several health benefits and is cost-effective in pediatric and adult patients with SCD [7-19]. Hydroxyurea induces fetal hemoglobin production, thereby decreasing sickle hemoglobin erythrocyte polymers, hemolysis, and vaso-occlusion [10,11]. Rigorous investigation over the past 30 years has demonstrated the efficacy of hydroxyurea in reducing disease complications, health care utilization, and costs for patients with SCD [7-18]. Consequently, the National Heart Lung and Blood Institute (NHLBI) has issued guidelines recommending its use among symptomatic adults and all children with SCD (HbSS and HbS$\beta^0$-thal genotypes) aged ≥9 months [19]. Despite being evidence-based and endorsed by NHLBI guidelines, adherence to hydroxyurea in SCD remains suboptimal [7,20-23]. Lower adherence rates have been associated with worse health outcomes, including more frequent SCD-related complications, low health-related quality of life, and increased health care utilization [4,7,15,24,25].

Medication adherence is a dynamic multifactorial process that reflects influential factors at three different levels: patient, health care provider, and health system [26]. In SCD, at the patient level, various barriers related to hydroxyurea adherence have been identified, such as concerns about efficacy and side effects, forgetfulness, inability to obtain refills, and lack of knowledge about hydroxyurea [20-23,25,27-33]. In addition, among patients with SCD, the number of adherence barriers is negatively correlated with the overall adherence rates to different medications, including hydroxyurea [20]. Optimizing hydroxyurea adherence throughout the lifespan is critical to improving health outcomes in this population, particularly among adolescents and young adults—a vulnerable subpopulation at a higher risk of low hydroxyurea adherence and disease-related morbidity and mortality [1-3,19,24,34,35]. In addition, adolescents and young adults are likely to start or are already managing their medications independently, making this an ideal developmental period to engage patients with SCD in building their self-management skills, including adherence to hydroxyurea, which is essential and could have long-term benefits [36-38]. Furthermore, achieving a better understanding of different patient barriers represents a key strategy for improving hydroxyurea adherence in SCD [20,23,25,31,33].

The Health Belief Model (HBM) is commonly used to explain the constructs underlying medication adherence behavior [39]. The HBM suggests that patients' adherence level is related to a number of factors that are independently weighed [39]. The health-related action driving the increased use of hydroxyurea includes 5 constructs within the HBM: perceived susceptibility, perceived severity, perceived benefits, perceived barriers, and self-efficacy. Notably, these 5 constructs represent modifiable factors that, together, can be influenced to increase the use of hydroxyurea.

Technological solutions are becoming more common in health care. Access to personal technology is nearly ubiquitous, and a growing number of patients use technology for their health care needs [40,41]. Mobile health (mHealth) interventions have been shown to improve patient activation and engagement [42-46], making them a possible tool to improve outcomes. In particular, earlier studies showed that individuals with SCD and their families were interested in using mHealth technologies to manage their health [22,47,48]. In other chronic health conditions (eg, asthma and diabetes mellitus), there is mounting evidence that self-care and self-management skills can be improved with the use of mHealth interventions [49-51]. Furthermore, recent systematic reviews showed promising data to support the overall feasibility, acceptability, and efficacy of mHealth interventions in promoting adherence behavior and improving health outcomes in different patient populations [52-57], including SCD [58]. In this study, we examined barriers and facilitators to hydroxyurea adherence informed by the HBM to support the development of a new mHealth intervention to foster greater hydroxyurea use.

### Study Objectives

To date, only a few reported mHealth interventions have focused on medication adherence among individuals with SCD [58-61]. Additionally, the integration and application of user-centered

design principles in the development of interventions have been limited within this population [58,62]. User-centered design has been defined as "an iterative design process in which designers focus on the users and their needs in each phase of the design process, with a call for involving users throughout the design process via a variety of research and design techniques so as to create highly usable and accessible products for them" [63]. The objectives of this study were to (1) use an agnostic approach to identify patient-level barriers to hydroxyurea adherence and utilization and (2) apply a user-centered design approach to develop a patient-informed SCD mobile app, *InCharge Health*, as a behavioral intervention to improve adherence to hydroxyurea. We hypothesized that hydroxyurea adherence barriers would vary among patients with SCD and that the barriers could be targeted by functionalities within an mHealth app.

## Methods

### Design

This study utilized user-centered design, which is an evidence-based and iterative approach that incorporates the needs and context of a specific end-user group (eg, adolescents and young adults with SCD) and helps ensure that the resulting digital health intervention is acceptable and effective [62,64]. The first step of this approach is to better understand the users' needs, context, and experiences [62,64]. Thus, phases 1 to 3 of this study utilized a mixed method approach, with qualitative data derived from an optimization digital innovation workshop (phase 1) and semistructured interviews (phase 3), and quantitative data derived from a needs assessment survey (phase 2). Together, these data were used to better understand the needs and experiences of individuals with SCD who aim to consistently take hydroxyurea and to inform the initial design of an app prototype. Consistent with the iterative nature of user-centered design, phase 4 of the study involved obtaining end-user feedback on the app prototype via focus groups to further tailor and inform the final design of the app.

### Participants

Individuals were eligible if they were between the ages of 15.0 and 44.9 years, had a diagnosis of SCD, and were affiliated with one of the multicenter National Institutes of Health (NIH)–funded SCD implementation consortium (SCDIC) sites. The development of *InCharge Health* is part of the multicenter NIH-funded SCDIC that aims at increasing the adoption of evidence-based interventions, including hydroxyurea treatment, among individuals aged 15.0 to 44.9 years with SCD using implementation science approaches [65].

Prior use of hydroxyurea was not an inclusion criterion for the study. Participants with an indication of hydroxyurea therapy, but no prior history of receiving this therapy, or who received this therapy but later discontinued it, were included as we also sought to obtain the perspectives of individuals who had previously been offered hydroxyurea but declined to take it, as such perspectives could provide further insight into potential barriers to taking hydroxyurea. Participants were recruited continuously from 2 SCDIC hematology clinics (St Jude Children's Research Hospital and Methodist University Hospital) and a community-based organization in Memphis, Tennessee, United States. Participants were recruited from these specific clinics as these were SCDIC sites. Recruitment took place during regular clinic visits. Participants received information about the study either via a flyer or via in-person communication with a study team member. All participants signed an informed consent before any study procedure. If participants were minors, their legal guardian signed the consent, and assent was obtained from the adolescent participants. The study was approved by the Institutional Review Boards from St Jude Children's Research Hospital and the Methodist Healthcare/University of Tennessee Health Science Center.

### Procedure

This study addressed the patient underutilization of hydroxyurea by developing a new mHealth intervention to improve hydroxyurea adherence. The study consisted of 4 phases, with a total of 118 participants. *Phase 1*: patients (n=6) and health care providers (n=12) participated in an optimization digital innovation workshop [62]. *Phase 2*: patients (n=99) completed surveys pertaining to demographic information, mHealth use, perceived barriers and facilitators to hydroxyurea use, and health literacy. *Phase 3*: patients (n=20) completed semistructured interviews pertaining to their experiences with apps and mobile phones, perceptions of hydroxyurea, perceptions of risk of SCD complications, and their SCD self-efficacy. *Phase 4*: patients (n=12) participated in focus groups to provide feedback on an initial app prototype. Feedback on the app and several of its proposed features were used to further refine the app prototype. Patients were compensated for their time in each phase of the study as follows: *Phase 1*, US $100; *Phase 2*, US $15; *Phase 3*, US $15; *Phase 4*, US $20. Completion of all study phases took 1 year in total.

#### *Phase 1: Optimization Digital Innovation Workshop*

Patients and providers were invited to participate in a 1-day workshop focused on identifying problems around hydroxyurea use and developing potential solutions. This workshop lasted 8 hours and included patients, clinicians, both hematologists and psychologists, researchers, and research staff. Preferences for features of the app were also identified through an iterative process and group discussions that started with the identification of gaps in hydroxyurea utilization (specifically why adherence was low) and ended with possible digital solutions to foster greater hydroxyurea use.

#### *Phase 2: Needs Assessment Survey*

This survey consisted of a total of 20 questions that aimed at investigating the barriers and facilitators to hydroxyurea use, patients' perception of the mHealth benefit, and the health literacy level of the population. The 4 main sections of the survey were as follows:

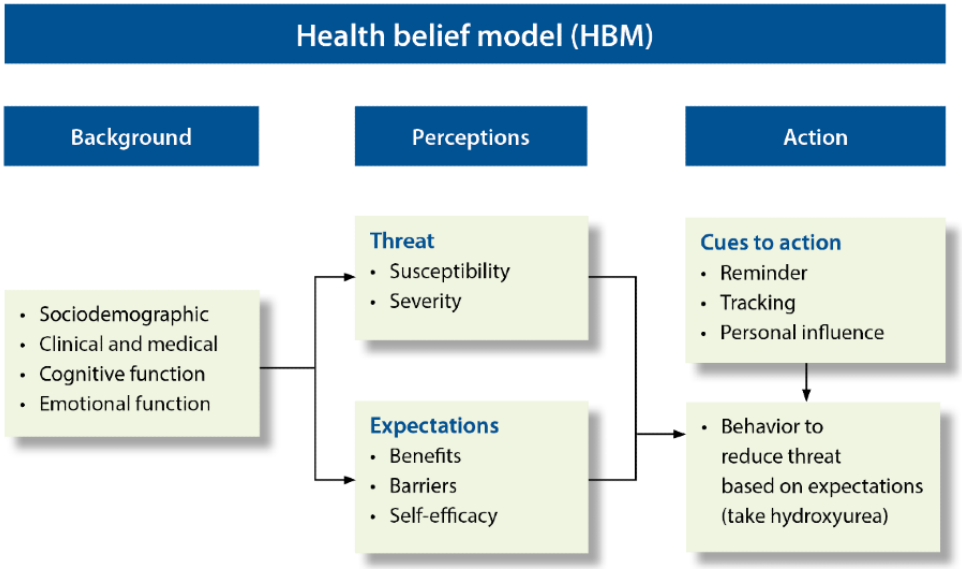
- Demographics: patients completed a demographic survey that assessed their age, gender, race, ethnicity, SCD type, zip code, household income, and household size.
- mHealth use: patients completed a 2-item study-specific survey assessing whether they are currently using or ever used mHealth to help them take their hydroxyurea.

- Hydroxyurea use and barriers: patients completed a 7-item study-specific survey assessing hydroxyurea use (eg, current and past use, if any, and duration if currently using hydroxyurea) and perceived barriers to hydroxyurea use.
- Health literacy: patients completed the Newest Vital Sign (NVS) [66], a 6-item validated screening tool that assesses an individual's ability to understand and apply information displayed on a nutrition label for ice cream. This tool measures both health literacy and numeracy.

### *Phase 3: Semistructured Interviews*

Individuals who agreed to participate in the survey were invited to take part in a 30 to 45-min semistructured interview that provided complementary data on the participant's knowledge and perceptions of hydroxyurea. Interview participants were also asked about their prior history with mHealth apps for medication adherence and perceptions. Interviews were conducted in the sickle cell clinic by a single study team member. The HBM [39] was used as a theoretical framework for the development of the final interview guide. Figure 1 illustrates the HBM as applied to adherence to hydroxyurea. Participants were asked specific questions relevant to the model, including questions about perceived barriers and benefits to taking hydroxyurea, perceived level of risk of having SCD complications, and self-efficacy regarding their ability to take care of their health.

**Figure 1.** Health belief model applied to hydroxyurea adherence.




### *Phase 4: Focus Groups*

Individuals who completed the surveys were asked to indicate their interest in participating in a focus group in the future. Those indicating an interest, and who had not already participated in the semistructured interviews, were invited to participate in the focus groups. The overall goal of the focus groups was to obtain feedback on the app prototype to assist with further app refinement.

### Data Synthesis and Analysis

Quantitative data were analyzed using SPSS version 25 (IBM, Armonk, New York). Descriptive and correlation analyses were used to summarize demographic and survey item data. With regard to the qualitative data obtained in phase 3 and phase 4, interviews and focus group discussions were audiotaped, transcribed, and entered into NVivo 12 for coding and analysis. Coding of the interview data was conducted by two members of the study team (JH and HK) following a cyclical coding process using a constructivist grounded theory approach [67-69]. A final coding scheme was developed using a consensus coding approach where study team members met weekly to discuss agreement and disagreement and rectified any discrepancies throughout the coding process [68]. The cyclical coding process involved 3 coding steps—Step 1: an initial set of codes on the basis of interview guide questions; Step 2: a subset of codes within Step 1 codes; Step 3: thematic codes identified by common codes found within Step 1 and 2 codes.

## Results

### Study Population

In phase 1**,** participants were 6 adult patients with SCD between the ages of 18 and 45 years (3 males and 3 females) and 13 health care providers, community advocates, and researchers, including 2 physicians, 1 nurse, 2 psychologists, 4 PhD researchers (behavioral and informatics), 1 health information security expert, 1 community-based organization member, and 2 research coordinators; 10 patients were not interested in participating in phase 1. Purposive sampling was used to recruit patient and provider/researcher participants to phase 1 of the study. Table 1 presents demographic data for patient participants in phases 2, 3, and 4. In phase 2, 100 participants were surveyed, and 99 patients completed the demographic survey, who were separated by age group in Multimedia Appendix 1. The upcoming clinic visits of potential patient participants were reviewed, and patients were randomly approached during their visits to participate in phase 2 of the study. Participant demographics were comparable to the larger SCD population [70]. In phase 2, 10 participants declined to participate; 12

declined to participate in phase 3, and 19 declined to participate in phase 4. The mean age of patients was 21.7 years (SD 6.4) and 46% (46/99) were female. Adolescents (aged 15-17.9 years) comprised 78% (78/99) of the total participants, and 21% (21/99) were between the ages of 18 and 44.9 years. Overall, 70% (70/99) of patients had HbSS or $S\beta^0$ thalassemia genotypes, and the remaining had HbSC, $HbS\beta^+$ thalassemia, or another variant. Almost all patients were African American (94/99, 94%), with the remaining 5.1% indicating "Other" and 1% choosing Hispanic or Latino as their ethnicity. For household income, 38% of participants either did not know or preferred not to answer, 47% had a household income of US $29,999 or lower, 9% indicated a household income between US $30,000 and US $59,999, and 6% were above US $60,000. These patients were randomly selected during their clinic visits, and their demographics were similar to the demographics of the longitudinal cohort study Sickle Cell Clinical Research and Intervention Program (SCCRIP), which encompasses nearly the entire population of the participating sites [70].

During clinic visits, participants who completed the survey were offered the opportunity to participate in the interviews until a sample size of 20 was reached. Thus, 20 adolescents and adults (average age 24.5 years, SD 9.28; 45% female) participated in *phase 3* of the study, which involved semistructured interviews. In *phase 4*, 12 adolescents and adults (average age 21.10 years, SD 6.01; 75% female) participated in focus groups where they were invited to review an interactive app prototype and provided feedback on functionality, key features (eg, medication reminders, tracking, and points/rewards), and overall design.

**Table 1.** Participant characteristics across phases 2, 3, and 4.

| Characteristics | Phase 2 enhanced NA[a] survey, (N=99) | Phase 3 interview, (N=20) | Phase 4 focus group, (N=12)[b] |
|---|---|---|---|
| Age (years), mean (SD) | 22 (6.46) | 24 (9.28) | 21 (6.01) |
| **Age group (years), n (%)** | | | |
| 15-17 | 31 (31) | 7 (35) | 5 (42) |
| 18-24 | 47 (48) | 7 (35) | 5 (42) |
| 25-34 | 9 (9) | 1 (5) | 1 (8) |
| 35-45 | 12 (12) | 5 (25) | 1 (8) |
| **Gender, n (%)** | | | |
| Female | 46 (46) | 9 (45) | 9 (75) |
| Male | 53 (53) | 11 (55) | 3 (25) |
| **Race n (%)** | | | |
| African American | 71 (96) | 19 (95) | 12 (100) |
| Other | 3 (4) | 1 (5) | —[c] |
| **SCD[d] genotype, n (%)** | | | |
| HbSS | 62 (63) | 12 (60) | 8 (67) |
| HbS$\beta^0$ thalassemia | 8 (8) | 1 (5) | — |
| HbSC | 23 (23) | 5 (25) | 3 (25) |
| HbS$\beta^+$ thalassemia | 4 (4) | 1 (5) | — |
| Other variant | 1 (1) | — | 1 (8) |
| Do not know | 1 (1) | 1 (5) | — |
| **Annual household income (US $), n (%)** | | | |
| Less than 5000 | 23 (23) | 2 (10) | 2 (17) |
| 5000-9999 | 6 (6) | 2 (10) | 1 (8) |
| 10,000-14,999 | 6 (6) | 2 (10) | — |
| 15,000-19,999 | 2 (2) | 1 (5) | — |
| 20,000-29,999 | 9 (9) | — | 1 (8) |
| 30,000-39,999 | — | — | — |
| 40,000-49,999 | 6 (6) | 1 (5) | 1 (8) |
| 50,000-59,999 | 3 (3) | 1 (5) | — |
| 60,000-79,999 | 3 (3) | — | 2 (17) |
| 80,000-94,999 | 2 (2) | — | — |
| 95,000 and above | 1 (1) | — | — |
| Prefer not to answer | 13 (13) | 1 (5) | 3 (25) |
| Do not know | 25 (25) | 10 (50) | 2 (17) |
| **NVS[e] health literacy, n (%)** | | | |
| High likelihood of limited literacy | 57 (58) | 11 (55) | 7 (58) |
| Possibility of limited literacy | 21 (21) | 6 (30) | 2 (17) |
| Adequate literacy | 21 (21) | 3 (15) | 3 (25) |

[a]NA: needs assessment.

[b]A proportion of patients participated in >1 study phase.

[c]No participants in this category.

[d]SCD: sickle cell disease.

[e]NVS: Newest Vital Sign.

## Phase 1

During the optimization digital innovation workshop (Multimedia Appendix 2), patients and providers discussed obstacles to taking hydroxyurea, which were then categorized and prioritized. Identified obstacles included (1) side effects, (2) quantity and size of the pills, (3) busy life and a tendency to forget medications, (4) cost of the medication, and (5) attitudes toward medication and negative emotions (eg, depression, "tired of taking medication," "no immediate or visible reward," stigma, anger, and hopelessness). Group discussion also focused on features to include in the app, which derived 7 categories: (1) gaming, competition, and rewards; (2) group support and community; (3) education, resources, and ability to contact a doctor/an expert; (4) reminders and notifications; (5) personalization; (6) visual checklists; and (7) graphical trackers and charts. Participants also reflected on the most important features of the app. Overall, features that were considered the most important included reminders to take hydroxyurea, followed by personalization and setting individual goals/motivations, motivation messages, gaming and points, tracking graphs, and finally features that allow communication with friends/family if the user forgets to take hydroxyurea.

## Phase 2

### *Mobile Health Use*

Overall, 24% (24/99) of those with a history of taking hydroxyurea had prior experience with text messages as an aid in taking hydroxyurea; 11% (11/99) indicated that they were currently using text messages to help them take hydroxyurea.

### *Hydroxyurea Use and Barriers*

Most patients (72/99, 73%) were taking hydroxyurea at the time of survey completion (current users), and 6% had taken hydroxyurea at some point in the past (previous users). Among those currently taking hydroxyurea, 43% (43/99) had taken it for >6 years, 42% (42/99) had taken it between 1 and 5 years, and 9% (9/99) had been on hydroxyurea for <1 year. Among patients who had any history with hydroxyurea, 79% (78/99) indicated "forgetting to take the medicine" and 24% (24/99) reported "it was hard to take the medicine at the right time" as the leading barriers to hydroxyurea adherence. The next most frequently endorsed barriers included "I don't like to think about having sickle cell disease when I am feeling well" (11/99, 11%) and "I am worried about side effects" (10/99, 10%). Of the total patients with a history of hydroxyurea use, 33% (33/99) indicated experiencing no barrier to hydroxyurea use. Facilitators to hydroxyurea use were rarely cited by the participants, who rather reported facilitators in the context of mHealth solutions.

### *Health Literacy*

Most patients were found to have either a high likelihood of limited health literacy (57/99, 58%) or possibly limited health literacy (21/99, 21%). Only 21% (21/99) of patients were found to have an adequate level of health literacy.

## Phase 3

Results from phase 3 are presented around the primary content areas explored during the semistructured interviews and in alignment with the core modifiable components of the HBM as applied to this study: background, perceptions pertaining to expectations, perceptions pertaining to threat, and cues to action.

### *App and Mobile Phone Experiences and Perceptions*

All participants indicated regular use of mobile phone apps for social media, communication with peers or family, and entertainment consumption (eg, to watch movies, play games, listen to music, or read books). A minority of participants stated experience with using an app to assist with health-related activities, such as using the phone's reminder function for medical appointments and medications or tracking physical activity. Overall, participants expressed positive perceptions about mobile phones and apps; however, when asked to state any concerns they might have, ensuring the protection of private data and general negativity in the realm of social media were the most common drawbacks to mobile phone apps.

### *Health Belief Model Perceptions: Benefits, Barriers, and Sickle Cell Disease Self-Efficacy*

#### Education

Overall, most participants did not perceive a need for additional education on hydroxyurea, although some did indicate additional education could be helpful, including information on the pros and cons of taking hydroxyurea.

#### Perceived Benefits of Taking Hydroxyurea

The primary perceived benefit of taking hydroxyurea was the improvement of overall health and quality of life. Participants specifically mentioned the decrease in pain crises and a decrease in hospitalizations as benefits of hydroxyurea.

#### Barriers to Taking Hydroxyurea

A large majority of the participants identified forgetting to take hydroxyurea as a primary barrier to adherence. Additional barriers included insurance or price, number of pills in the dose, fatigue, competing activities (eg, busy schedule), and feeling "tired" of taking medications. Participants who were not currently taking hydroxyurea described several barriers to beginning or restarting hydroxyurea including hearing about someone who experienced side effects, currently taking several medications, never having been offered hydroxyurea by a physician, personal preference, and a lack of need to take it because of perceived low SCD severity.

#### High Sickle Cell Disease Self-Efficacy

Overall, the vast majority of participants described feeling confident or very confident about their ability to take care of their health, attend their medical appointments, and take hydroxyurea. Most participants also rated this confidence as 8 or 10 out of 10 (10 being the most confident).

### *Health Belief Model Perceptions: Susceptibility to Sickle Cell Disease Complications, Perceived Severity, and Sickle Cell Disease–Related Worry*

#### Low Perceived Risk

Overall, most participants viewed themselves as at low risk of SCD complications. When asked why they are at low risk, participants reported engaging in strategies such as taking their medications, drinking water, dressing for the weather, and abstaining from risky health behaviors (eg, drinking alcohol and smoking). For those who saw themselves as being at high risk, reasons for this perception included experiencing elevated pain and an increased number of surgeries.

#### Increased Worry About Complications

Interestingly, although most participants perceived themselves as being at low risk for SCD complications, most also reported experiencing worry about SCD complications, maintaining good health, and negative health-related outcomes such as hospitalization, blood clots or stroke, and the cumulative damage of SCD. The occurrence of pain and pain crises was the most common worry reported by participants. It should also be noted that a subset of participants denied experiencing worry about SCD complications.

### *Health Belief Model Cues to Action: Preferred Features in an App for Hydroxyurea Adherence*

#### Trackers and Reminders

Several participants indicated that they would like to have reminders within the app, including the ability to customize these reminders on the basis of their personal preferences. Participants also indicated it would be helpful to have the ability to track their adherence to hydroxyurea and other medications.

#### Education and Communication

Some participants also expressed a desire for the app to provide additional information about hydroxyurea (eg, dosage), and some expressed a desire to connect with other patients and friends and family (eg, parents) to increase accountability for taking hydroxyurea. Some participants also described a desire to communicate with their medical team via the app.

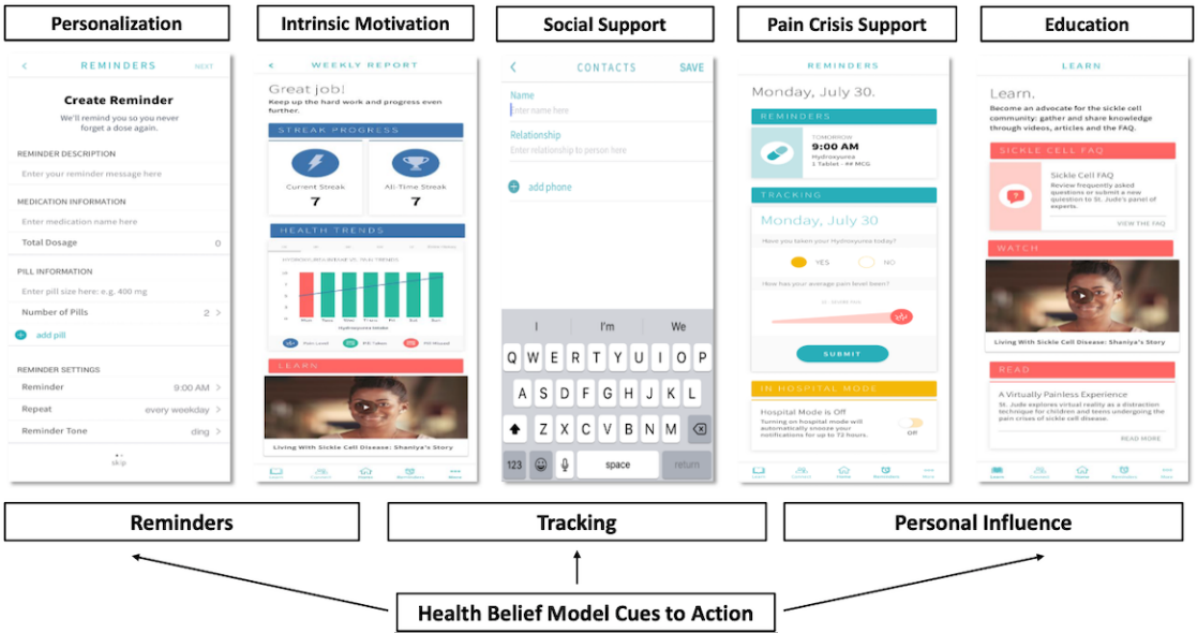
#### Engaging and User-Friendly

Overall, there was a general desire to have an app that was easy to use while also engaging. In being able to customize reminders, several participants reported that they would like to have the ability to make unique and exciting medication reminders linked to their personal motivation for staying healthy.

## Phase 4

Data derived from the focus groups revealed 5 primary themes: (1) personalization, (2) intrinsic motivation, (3) social support, (4) support during pain crisis, and (5) education. Participants' input from this phase led to the development of the *InCharge Health* app (Figure 2).

**Figure 2.** Screenshots of different functions in InCharge Health App mapped to health belief model.




### *Personalization*

Participants expressed a desire to personalize various aspects of the app, including reminders, motivational messages, and educational content. Among adolescents, in particular, being able to put reminders in their own words was viewed as making the app feel less like "nagging." Overall, participants also indicated that they would be less likely to ignore reminders if they were in their own words. As a result of this feedback, within the *InCharge Health* app, patients have the option of customizing the reminder description and the day and time that they receive the reminder. The customized reminder feature is shown in Figure 2.

### *Intrinsic Motivation*

Contrasting views with regard to motivation to use the app were found among adolescents and adults. Specifically, adolescents liked the idea of short-term rewards being offered to increase motivation, whereas adults generally held the view that motivation should be more intrinsically based and come from a personal desire to support health and well-being. To accommodate both views, the concept of *streaks* (see Figure 2) was utilized within the app, such that participants can earn streaks when they take their hydroxyurea several days in a row and can, therefore, track their progress. In addition, participants can rate their pain each day and are provided with a weekly tracking report of their pain, which is, in turn, graphed next to their adherence to hydroxyurea (Figure 2). The rationale behind this feature is that directly showing patients changes in pain over time and how these are connected to changes in adherence to hydroxyurea will potentially increase patient motivation to take hydroxyurea consistently.

### *Social Support*

Participants also expressed a desire to connect with others with shared experiences. Many reported already connecting with others with SCD via social media and online forums and liked the ability to connect with peers and experts all in one place. In response to this feedback, on the homepage of the *InChargeHealth* app, patients have the option of connecting with their doctor, their medical chart, and other patients via the website, *one SCDvoice* [71]. They also have the option of adding *accountability contacts* who are sent messages when the patient does not indicate taking their medication for >4 hours (Figure 2).

### *Pain Crisis Support*

Participants also expressed an interest in app features that would help provide support during a pain crisis. In response to this feedback, patients can monitor their pain across time by recording their daily pain level within the app (Figure 2). In addition, if a pain crisis occurs, patients can contact their clinic directly from within the app, access educational information about pain treatment, and talk to other patients about their experiences with pain.

### *Education*

Participants also expressed interest in having access to educational content—both for themselves and for friends and family members. Specifically, participants were interested in content related to how hydroxyurea works, side effects, and pain management and tracking. In response, the *InCharge Health* app includes educational content pertaining to SCD, hydroxyurea, and more general health topics (eg, depression; Figure 2). Moreover, given the low levels of health literacy observed in phase 2, education and information more generally were provided in the app via several visual prompts (eg, graphs and images) and multiple formats, including videos, written materials (PDF files), and links to other websites. Additional app features accounting for low levels of health literacy observed included easy-to-read and simple instructions and links to the electronic medical record and the clinic to assist with navigation of the health system.

The HBM components of perceived barriers, cues to action, and self-efficacy emerged in the survey, semistructured interview, and focus group results and were subsequently incorporated in the development of the *InCharge Health* app (Figure 2). Forgetting to take medication was the largest reported barrier to hydroxyurea adherence. The app addresses this with cues to action with text messaging reminders that can be personalized by the user, and the user can enter accountability contacts who are notified if the user does not take their medication within the specified time period as indicated through the app. Participants also reported concern about medication side effects as a barrier; therefore, the app provides educational content through readings and videos. Additionally, the app tracks consistent hydroxyurea use through *streaks*, and users can track their pain experience and see how pain and medication use are connected, which in turn can increase intrinsic motivation and self-efficacy.

## Discussion

### Principal Findings

Hydroxyurea is one of only 2 FDA-approved medications for SCD, and the one with the most amount of evidence for its efficacy. Despite its evidence for clinical benefit, patients with SCD face difficulty in maintaining adequate adherence. Using a user-centered approach that investigated the reasons for poor adherence on the basis of the HBM, we identified the important drivers of poor adherence and translated them into an mHealth solution that addresses all important barriers, the *InCharge Health* app.

The results of this study indicated that there is an inadequate level of health literacy among adolescents and young adults with SCD, which is consistent with prior research showing that health literacy is suboptimal among adolescents [72,73] and adults with SCD as well as caregivers [73]. Prior research in other chronic diseases has also shown low health literacy to be related to poor medication adherence [74], thus further highlighting the need to provide appropriate and tailored education to patients with SCD who are attempting to consistently take hydroxyurea. Importantly, the *InCharge Health* app was developed accounting for the limited health literacy level of our population. Specifically, text information within the app was purposely written in a layperson format and at a sixth-grade reading level to increase the patients' ability to understand the information provided. In addition, education and information more generally were provided via several visual prompts and multiple formats. In sum, these multi-component features are simple, intuitive, and aim to provide easy-to-understand information to patients on their condition and effective treatments. Furthermore, as low health literacy is likely strongly associated with both education level [72] and cognitive functioning among individuals with SCD, our findings have implications for future investigations examining the influence of socioeconomic factors and executive functioning among youths and adults with SCD. Within the SCCRIP study [70], we examined the relationship between health literacy and executive functioning and educational level. Results obtained from these studies will assist in informing future interventions targeting health literacy.

### Barriers to Hydroxyurea Adherence

SCD patients likely engage in an implicit assessment or an individual cost-benefit evaluation where possible benefits from hydroxyurea are weighed against concerns about using it [75-77]. Our findings showed that a significant number of our adolescents and adults with SCD had different positive and negative views about hydroxyurea, which could affect their perceptions of the necessity of using it, which was consistent with earlier reports in SCD [20,23,27,55,78] and other chronic medical conditions [75,77,79-86]. In a study by Haywood et al [27], SCD patients reported a number of concerns related to the use of hydroxyurea, including insufficient knowledge, lack of perceived benefits, and possible side effects, similar to our findings. In support of our findings, forgetfulness has been a common reason for low medication adherence among individuals with SCD and other patient populations [20,26,87]. In particular, among patients with SCD, forgetfulness could be exacerbated by undiagnosed or underestimated poor executive functioning and other cognitive deficits experienced in this population [88-90]. Consistent with our findings, earlier studies have reported access barriers and difficulties obtaining refills as a major challenge to hydroxyurea adherence [20,21]. Therefore, understanding patients' views and perceptions of hydroxyurea and possible access barriers is critical to developing an intervention that is customized to a particular population, increasing its likelihood for adoption and ability to improve adherence levels. We believe *InCharge Health* addresses these important barriers identified in our population and is supported by the literature.

Overall, the results of this study are consistent with prior literature examining barriers to medication adherence and the use of mHealth among diverse and minority populations. For example, in a cohort of urban minority adults with chronic obstructive pulmonary disease, nonadherence to medications was associated with low income, fewer years of informal education, and concerns about medication [91]. These results emphasize the importance of our findings pertaining to suboptimal levels of low health literacy among individuals with SCD and the potential impact on medication adherence and the tailoring of interventions to account for this factor. Within this study, the worry about side effects of hydroxyurea was an identified barrier to adherence, which is consistent with research among low-income, racially diverse adults with type 2 diabetes, which indicated that a common barrier to adherence among these adults was the belief that medications are harmful [92]. Interestingly, forgetting to take the medication was not a common barrier among this sample of adults with type 2 diabetes, unlike this study, where it was the most common barrier reported by patients with SCD. This is perhaps partially indicative of the cognitive effects prominent within SCD [93-95] versus other diseases such as diabetes.

### Mobile Health Solutions to Improve Medication Adherence in Sickle Cell Disease

Recent data reported high access to personal technology, including mobile devices, desktops, laptops, tablets, or iPads, in the general population and among adolescents and adults with SCD [22,41,48,96]. Our study focused on adolescents and adults with SCD who provided invaluable insights into the most suitable platform for technology-based interventions in our population. mHealth interventions represent a promising approach for improving hydroxyurea adherence in patients with SCD, including text messaging or directly observed therapy using mobile phones [59-61,97,98]. In our study, we also evaluated participants' preferences for a mobile phone app that would help to promote hydroxyurea adherence. Our participants identified medication reminder/tracker, education, communication, personalization, motivation, support during the pain episode, and social support as the most important features for their SCD-specific app. Our findings were consistent with our preliminary data that reported similar mobile phone app preferences among adolescents and young adults with SCD [22]. However, a unique feature of our study, and contribution to this body of work, is our inclusion of both adolescents and adults (aged 18-45 years) and taking a user-centered approach from concept elicitation to mobile phone app development (*InCharge Health*). In addition, our study was informed by an established theoretical model, the HBM [39], at different phases of the project. Recent evidence from several systematic reviews of medication adherence interventions [53-56,58,99,100] and experience in patients with SCD [22,58], cystic fibrosis [101], diabetes [102], and chronic pain [103] suggest that a personalized user-centered approach with a multi-functional mobile phone app may have the potential to improve low hydroxyurea adherence among adults with SCD. This approach is more likely to help overcome adherence barriers and engage participants with the app over time.

Although the use of mHealth among diverse and minority populations has often been an understudied area within the broader digital health intervention literature, this study adds to a quickly evolving field that places increased emphasis on examinations of acceptability and efficacy of mHealth interventions for underserved and minority populations. For example, there have been recent examinations of mHealth interventions (eg, text message, app, and social media based) to increase medication adherence among HIV-positive men who have sex with men [104], of app-based and culturally tailored self-management programs targeting blood pressure among hypertensive Hispanic adults [105], and of a text messaging–based intervention designed to increase medication adherence among low-income, diverse adults with type 2 diabetes [92]. Recent work has also provided evidence for the acceptability of using an app-based assessment of breast cancer risk among ethnically diverse, older, and low-income women [106] and the reliability and validity of mobile phone-based self-monitoring among African American and Latina mothers [107]. Despite this progress, it is still important to note that although the number of individuals with access to the internet and mobile phones continues to rise, varying levels of access and mHealth usage still exist among vulnerable populations nationally and internationally [58,108-111].

### Limitations

Our study had a few limitations that are worth noting. First, data were collected from patients treated at academic institutions and members of a community-based organization, all in Memphis, as a convenience sample of adults with SCD, which

might limit the generalizability of our results to other sickle cell centers. Second, selection bias can be a concern where motivated patients are more likely to participate in clinical research studies. However, our interviews were intentional at enrolling patients during their regular clinic visits, potentially reducing a selection of those most motivated to attend a separate research visit. Patients with SCD who were not compliant with their clinic visits or did not participate in the study for different reasons (eg, refusal) may have provided different insights and additional information to inform app development. Despite the potential selection of motivated participants, our results on mobile phone app preferences were consistent with published studies [22,48]. Third, in addition to the NVS, which is a validated tool that was tested in SCD, we used another nonvalidated survey instrument to examine hydroxyurea barriers and facilitators. Nevertheless, our survey was informed by an established behavioral theory (ie, HBM) and developed by experts in the field. Fourth, the wide age range within this study somewhat limits the extent to which the app design and content can be tailored to fit the needs and preferences of individuals with SCD of all ages. Finally, we did not examine patients' access to different personal technology tools (eg, iPhone or Android mobile phones, tablets, or laptops). However, prior studies reported wide access to these devices among adolescents and adults with SCD [22,47,48], similar to the general population [40,41].

## Conclusions

In conclusion, patients with SCD and health care providers identified several challenges with optimal hydroxyurea adherence. Our findings of barriers to hydroxyurea utilization replicated those in the literature, utilized the HBM as the conceptual framework for understanding the health behavior of hydroxyurea adherence, and directly informed the translation of mHealth into an intervention to modify hydroxyurea adherence. The invaluable insights of patients with SCD on several preferred app features guided, in all stages of the design, the development of a novel adherence-enhancing multicomponent mobile phone app, *InCharge Health*, focused on hydroxyurea adherence. The efficacy and implementation of *InCharge Health* as an mHealth intervention to promote hydroxyurea adherence will be tested in a multicenter trial (NCT04080167) [112]. If this new mHealth intervention is proven efficacious, it can be incorporated into the regular care of patients with SCD as an adjuvant to improve medication adherence. Given the existing evidence, mHealth interventions have considerable potential to optimize medication adherence, quality of life, and health outcomes in patients with SCD and other chronic health conditions.

## Acknowledgments

This project was supported by grant number 1U01HL133996-01 (principal investigator [PI]: JH) from the NIH, NHLBI; K12HS023011 (PI: SB) from the Agency for Healthcare Research and Quality (AHRQ); and K01 HL125495 (PI: JP) from the NIH, NHLBI. The content is solely the responsibility of the authors and does not necessarily represent the official views of AHRQ or NIH/NHLBI. The authors would like to thank Yvonne Carroll, JD, RN for helpful suggestions during development of *InCharge Health*, Mr Michael A Jackson (Sickle Cell Foundation of Tennessee) for contributing with app content selection, Jeff Frey, PhD, MBA, BS, for his assistance in organizing and leading the optimization digital workshop, and Agency 39A, LLC [113] for developing the *InCharge Health* app prototype.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Phase 2 participant characteristics by age group (N=99).
[DOCX File , 26 KB - mhealth_v8i5e14884_app1.docx ]

Multimedia Appendix 2
Optimization digital innovation workshop for hydroxyurea.
[PDF File (Adobe PDF File), 4126 KB - mhealth_v8i5e14884_app2.pdf ]

## References

1. Hassell KL. Population estimates of sickle cell disease in the US. Am J Prev Med 2010 Apr;38(4 Suppl):S512-S521. [doi: 10.1016/j.amepre.2009.12.022] [Medline: 20331952]
2. Centers for Disease Control and Prevention (CDC). Sickle Cell Disease: Data & Statistics on Sickle Cell Disease URL: http://www.cdc.gov/ncbddd/sicklecell/data.html [accessed 2018-10-14]
3. Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. Lancet 2010 Dec 11;376(9757):2018-2031. [doi: 10.1016/S0140-6736(10)61029-X] [Medline: 21131035]
4. Panepinto JA, Bonner M. Health-related quality of life in sickle cell disease: past, present, and future. Pediatr Blood Cancer 2012 Aug;59(2):377-385. [doi: 10.1002/pbc.24176] [Medline: 22522407]

5. Boulet SL, Yanni EA, Creary MS, Olney RS. Health status and healthcare use in a national sample of children with sickle cell disease. Am J Prev Med 2010 Apr;38(4 Suppl):S528-S535. [doi: 10.1016/j.amepre.2010.01.003] [Medline: 20331954]
6. Lanzkron S, Carroll CP, Haywood C. The burden of emergency department use for sickle-cell disease: an analysis of the national emergency department sample database. Am J Hematol 2010 Oct;85(10):797-799 [FREE Full text] [doi: 10.1002/ajh.21807] [Medline: 20730795]
7. Candrilli SD, O'Brien SH, Ware RE, Nahata MC, Seiber EE, Balkrishnan R. Hydroxyurea adherence and associated outcomes among Medicaid enrollees with sickle cell disease. Am J Hematol 2011 Mar;86(3):273-277 [FREE Full text] [doi: 10.1002/ajh.21968] [Medline: 21328441]
8. Thornburg CD, Calatroni A, Panepinto JA. Differences in health-related quality of life in children with sickle cell disease receiving hydroxyurea. J Pediatr Hematol Oncol 2011 May;33(4):251-254 [FREE Full text] [doi: 10.1097/MPH.0b013e3182114c54] [Medline: 21516020]
9. Wang WC, Oyeku SO, Luo Z, Boulet SL, Miller ST, Casella JF, BABY HUG Investigators. Hydroxyurea is associated with lower costs of care of young children with sickle cell anemia. Pediatrics 2013 Oct;132(4):677-683 [FREE Full text] [doi: 10.1542/peds.2013-0333] [Medline: 23999955]
10. Wang W, Ware RE, Miller ST, Iyer RV, Casella JF, Minniti CP, BABY HUG investigators. Hydroxycarbamide in very young children with sickle-cell anaemia: a multicentre, randomised, controlled trial (BABY HUG). Lancet 2011 May 14;377(9778):1663-1672 [FREE Full text] [doi: 10.1016/S0140-6736(11)60355-3] [Medline: 21571150]
11. Charache S, Terrin ML, Moore RD, Dover GJ, Barton FB, Eckert SV, et al. Effect of hydroxyurea on the frequency of painful crises in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. N Engl J Med 1995 May 18;332(20):1317-1322. [doi: 10.1056/NEJM199505183322001] [Medline: 7715639]
12. Thornburg CD, Files B, Luo Z, Miller ST, Kalpatthi R, Iyer R, BABY HUG Investigators. Impact of hydroxyurea on clinical events in the BABY HUG trial. Blood 2012 Nov 22;120(22):4304-10; quiz 4448 [FREE Full text] [doi: 10.1182/blood-2012-03-419879] [Medline: 22915643]
13. Moore RD, Charache S, Terrin ML, Barton FB, Ballas SK. Cost-effectiveness of hydroxyurea in sickle cell anemia. Investigators of the Multicenter Study of Hydroxyurea in Sickle Cell Anemia. Am J Hematol 2000 May;64(1):26-31 [FREE Full text] [doi: 10.1002/(sici)1096-8652(200005)64:1<26::aid-ajh5>3.0.co;2-f] [Medline: 10815784]
14. Ballas SK, Barton FB, Waclawiw MA, Swerdlow P, Eckman JR, Pegelow CH, et al. Hydroxyurea and sickle cell anemia: effect on quality of life. Health Qual Life Outcomes 2006 Aug 31;4:59 [FREE Full text] [doi: 10.1186/1477-7525-4-59] [Medline: 16942629]
15. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Health-related quality of life and adherence to hydroxyurea in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2017 Jun;64(6). [doi: 10.1002/pbc.26369] [Medline: 27896936]
16. Hankins J, Ware R, Rogers Z, Wynn LW, Lane PA, Scott JP, et al. Long-term hydroxyurea therapy for infants with sickle cell anemia: the HUSOFT extension study. Blood 2005 Oct 1;106(7):2269-2275 [FREE Full text] [doi: 10.1182/Blood-2004-12-4973] [Medline: 16172253]
17. Aygun B, Mortier NA, Smeltzer MP, Shulkin BL, Hankins JS, Ware RE. Hydroxyurea treatment decreases glomerular hyperfiltration in children with sickle cell anemia. Am J Hematol 2013 Feb;88(2):116-119 [FREE Full text] [doi: 10.1002/ajh.23365] [Medline: 23255310]
18. Estepp JH, Smeltzer MP, Wang WC, Hoehn ME, Hankins JS, Aygun B. Protection from sickle cell retinopathy is associated with elevated HbF levels and hydroxycarbamide use in children. Br J Haematol 2013 May;161(3):402-405. [doi: 10.1111/bjh.12238] [Medline: 23384083]
19. Yawn BP, Buchanan GR, Afenyi-Annan AN, Ballas SK, Hassell KL, James AH, et al. Management of sickle cell disease: summary of the 2014 evidence-based report by expert panel members. J Am Med Assoc 2014 Sep 10;312(10):1033-1048. [doi: 10.1001/jama.2014.10517] [Medline: 25203083]
20. Loiselle K, Lee JL, Szulczewski L, Drake S, Crosby LE, Pai AL. Systematic and meta-analytic review: medication adherence among pediatric patients with sickle cell disease. J Pediatr Psychol 2016 May;41(4):406-418 [FREE Full text] [doi: 10.1093/jpepsy/jsv084] [Medline: 26384715]
21. Thornburg CD, Calatroni A, Telen M, Kemper AR. Adherence to hydroxyurea therapy in children with sickle cell anemia. J Pediatr 2010 Mar;156(3):415-419 [FREE Full text] [doi: 10.1016/j.jpeds.2009.09.044] [Medline: 19880135]
22. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
23. Walsh KE, Cutrona SL, Kavanagh PL, Crosby LE, Malone C, Lobner K, et al. Medication adherence among pediatric patients with sickle cell disease: a systematic review. Pediatrics 2014 Dec;134(6):1175-1183 [FREE Full text] [doi: 10.1542/peds.2014-0177] [Medline: 25404717]
24. Badawy SM, Thompson AA, Holl JL, Penedo FJ, Liem RI. Healthcare utilization and hydroxyurea adherence in youth with sickle cell disease. Pediatr Hematol Oncol 2018;35(5-6):297-308. [doi: 10.1080/08880018.2018.1505988] [Medline: 30636474]

25. Badawy SM, Thompson AA, Lai J, Penedo FJ, Rychlik K, Liem RI. Adherence to hydroxyurea, health-related quality of life domains, and patients' perceptions of sickle cell disease and hydroxyurea: a cross-sectional study in adolescents and young adults. Health Qual Life Outcomes 2017 Jul 5;15(1):136 [FREE Full text] [doi: 10.1186/s12955-017-0713-x] [Medline: 28679417]
26. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
27. Haywood C, Beach MC, Bediako S, Carroll CP, Lattimer L, Jarrett D, et al. Examining the characteristics and beliefs of hydroxyurea users and nonusers among adults with sickle cell disease. Am J Hematol 2011 Jan;86(1):85-87 [FREE Full text] [doi: 10.1002/ajh.21883] [Medline: 21117058]
28. Brandow AM, Panepinto JA. Monitoring toxicity, impact, and adherence of hydroxyurea in children with sickle cell disease. Am J Hematol 2011 Sep;86(9):804-806 [FREE Full text] [doi: 10.1002/ajh.22101] [Medline: 21815183]
29. Oyeku SO, Driscoll MC, Cohen HW, Trachtman R, Pashankar F, Mullen C, et al. Parental and other factors associated with hydroxyurea use for pediatric sickle cell disease. Pediatr Blood Cancer 2013 Apr;60(4):653-658 [FREE Full text] [doi: 10.1002/pbc.24381] [Medline: 23129068]
30. Modi AC, Crosby LE, Guilfoyle SM, Lemanek KL, Witherspoon D, Mitchell MJ. Barriers to treatment adherence for pediatric patients with sickle cell disease and their families. Child Health Care 2009;38(2):107-122. [doi: 10.1080/02739610902813245]
31. Badawy SM, Thompson AA, Penedo FJ, Lai J, Rychlik K, Liem RI. Barriers to hydroxyurea adherence and health-related quality of life in adolescents and young adults with sickle cell disease. Eur J Haematol 2017 Jun;98(6):608-614. [doi: 10.1111/ejh.12878] [Medline: 28306171]
32. Badawy SM, Barrera L, Cai S, Thompson AA. Association between participants' characteristics, patient-reported outcomes, and clinical outcomes in youth with sickle cell disease. Biomed Res Int 2018;2018:8296139 [FREE Full text] [doi: 10.1155/2018/8296139] [Medline: 30105252]
33. Badawy SM, Thompson AA, Liem RI. Beliefs about hydroxyurea in youth with sickle cell disease. Hematol Oncol Stem Cell Ther 2018 Sep;11(3):142-148 [FREE Full text] [doi: 10.1016/j.hemonc.2018.01.001] [Medline: 29397333]
34. Kayle M, Docherty SL, Sloane R, Tanabe P, Maslow G, Pan W, et al. Transition to adult care in sickle cell disease: A longitudinal study of clinical characteristics and disease severity. Pediatr Blood Cancer 2019 Jan;66(1):e27463 [FREE Full text] [doi: 10.1002/pbc.27463] [Medline: 30251318]
35. Brousseau D, Panepinto J, Nimmer M, Hoffmann R. The number of people with sickle-cell disease in the United States: national and state estimates. Am J Hematol 2010 Jan;85(1):77-78 [FREE Full text] [doi: 10.1002/ajh.21570] [Medline: 20029951]
36. Staples B, Bravender T. Drug compliance in adolescents: assessing and managing modifiable risk factors. Paediatr Drugs 2002;4(8):503-513. [doi: 10.2165/00128072-200204080-00003] [Medline: 12126454]
37. Smith BA, Shuchman M. Problem of nonadherence in chronically ill adolescents: strategies for assessment and intervention. Curr Opin Pediatr 2005 Oct;17(5):613-618. [doi: 10.1097/01.mop.0000176443.26872.6e] [Medline: 16160536]
38. KyngAs H, Kroll T, Duffy M. Compliance in adolescents with chronic diseases: a review. J Adolesc Health 2000 Jun;26(6):379-388. [doi: 10.1016/s1054-139x(99)00042-7] [Medline: 10822178]
39. Stretcher V, Rosenstock IM. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, and Practice. San Francisco: Jossey-Bass; 1997.
40. Fox S. Pew Internet Research. 2014 Feb 13. Health Information Online URL: http://www.pewinternet.org/2014/02/13/health-information-online-2/ [accessed 2019-04-03]
41. Pew Research Center. 2019 Jun 12. Mobile Fact Sheet URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2019-04-03]
42. Haas K, Martin A, Park K. Text message intervention (TEACH) improves quality of life and patient activation in Celiac disease: a randomized clinical trial. J Pediatr 2017 Jun;185:62-7.e2 [FREE Full text] [doi: 10.1016/j.jpeds.2017.02.062] [Medline: 28343658]
43. John M, Samson-Akpan P, Etowa J, Akpabio I, John E. Enhancing self-care, adjustment and engagement through mobile phones in youth with HIV. Int Nurs Rev 2016 Dec;63(4):555-561. [doi: 10.1111/inr.12313] [Medline: 27653013]
44. Knoerl R, Lee D, Yang J, Bridges C, Kanzawa-Lee G, Smith GL, et al. Examining the impact of a web-based intervention to promote patient activation in chemotherapy-induced peripheral neuropathy assessment and management. J Cancer Educ 2018 Oct;33(5):1027-1035. [doi: 10.1007/s13187-017-1200-0] [Medline: 28265863]
45. Milani RV, Lavie CJ, Bober RM, Milani AR, Ventura HO. Improving hypertension control and patient engagement using digital tools. Am J Med 2017 Jan;130(1):14-20. [doi: 10.1016/j.amjmed.2016.07.029] [Medline: 27591179]
46. Solomon M, Wagner SL, Goes J. Effects of a web-based intervention for adults with chronic conditions on patient activation: online randomized controlled trial. J Med Internet Res 2012 Feb 21;14(1):e32 [FREE Full text] [doi: 10.2196/jmir.1924] [Medline: 22353433]
47. Cronin RM, Hankins JS, Adams-Graves P, Thompson AA, Kalinyak K, Byrd J, et al. Barriers and facilitators to research participation among adults, and parents of children with sickle cell disease: A trans-regional survey. Am J Hematol 2016 Oct;91(10):E461-E462 [FREE Full text] [doi: 10.1002/ajh.24483] [Medline: 27421796]

48. Shah N, Jonassaint J, De Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2014;38(2):99-103. [doi: 10.3109/03630269.2014.880716] [Medline: 24512633]
49. Al-Durra M, Torio M, Cafazzo JA. The use of behavior change theory in internet-based asthma self-management interventions: a systematic review. J Med Internet Res 2015 Apr 2;17(4):e89 [FREE Full text] [doi: 10.2196/jmir.4110] [Medline: 25835564]
50. de Jongh T, Gurol-Urganci I, Vodopivec-Jamsek V, Car J, Atun R. Mobile phone messaging for facilitating self-management of long-term illnesses. Cochrane Database Syst Rev 2012 Dec 12;12:CD007459 [FREE Full text] [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
51. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
52. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
53. 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]
54. Majeed-Ariss R, Baildam E, Campbell M, Chieng A, Fallon D, Hall A, et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic or long-term physical conditions. J Med Internet Res 2015 Dec 23;17(12):e287 [FREE Full text] [doi: 10.2196/jmir.5043] [Medline: 26701961]
55. Badawy SM, Kuhns LM. Texting and mobile phone app interventions for improving adherence to preventive behavior in adolescents: a systematic review. JMIR Mhealth Uhealth 2017 Apr 19;5(4):e50 [FREE Full text] [doi: 10.2196/mhealth.6837] [Medline: 28428157]
56. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
57. Badawy SM, Thompson AA, Kuhns LM. Medication adherence and technology-based interventions for adolescents with chronic health conditions: a few key considerations. JMIR Mhealth Uhealth 2017 Dec 22;5(12):e202 [FREE Full text] [doi: 10.2196/mhealth.8310] [Medline: 29273573]
58. Badawy SM, Cronin RM, Hankins J, Crosby L, DeBaun M, Thompson AA, et al. Patient-centered eHealth interventions for children, adolescents, and adults with sickle cell disease: systematic review. J Med Internet Res 2018 Jul 19;20(7):e10940 [FREE Full text] [doi: 10.2196/10940] [Medline: 30026178]
59. Creary SE, Gladwin MT, Byrne M, Hildesheim M, Krishnamurti L. A pilot study of electronic directly observed therapy to improve hydroxyurea adherence in pediatric patients with sickle-cell disease. Pediatr Blood Cancer 2014 Jun;61(6):1068-1073. [doi: 10.1002/pbc.24931] [Medline: 24436121]
60. Estepp JH, Winter B, Johnson M, Smeltzer MP, Howard SC, Hankins JS. Improved hydroxyurea effect with the use of text messaging in children with sickle cell anemia. Pediatr Blood Cancer 2014 Nov;61(11):2031-2036. [doi: 10.1002/pbc.25177] [Medline: 25132074]
61. Creary S, Chisolm D, Stanek J, Hankins J, O'Brien SH. A multidimensional electronic hydroxyurea adherence intervention for children with sickle cell disease: single-arm before-after study. JMIR Mhealth Uhealth 2019 Aug 8;7(8):e13452 [FREE Full text] [doi: 10.2196/13452] [Medline: 31397291]
62. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
63. Interaction Design Foundation. User Centered Design URL: https://www.interaction-design.org/literature/topics/user-centered-design [accessed 2019-06-16]
64. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
65. DiMartino LD, Baumann AA, Hsu LL, Kanter J, Gordeuk VR, Glassberg J, Sickle Cell Disease Implementation Consortium. The sickle cell disease implementation consortium: translating evidence-based guidelines into practice for sickle cell disease. Am J Hematol 2018 Dec;93(12):E391-E395 [FREE Full text] [doi: 10.1002/ajh.25282] [Medline: 30203558]
66. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
67. Thornberg R, Perhamus L, Charmaz K. Grounded Theory. Contemp Perspect Ear 2015:405-439 [FREE Full text] [doi: 10.1002/9781405165518.wbeosg070.pub2]
68. Saldana J. The Coding Manual For Qualitative Researchers. Third Edition. Thousand Oaks, CA: Sage Publications Ltd; 2016.
69. Charmaz K. Constructing Grounded Theory. Second Edition. Thousand Oaks, CA: Sage Publications Ltd; 2014.

70. Hankins JS, Estepp JH, Hodges JR, Villavicencio MA, Robison LL, Weiss MJ, et al. Sickle Cell Clinical Research and Intervention Program (SCCRIP): A lifespan cohort study for sickle cell disease progression from the pediatric stage into adulthood. Pediatr Blood Cancer 2018 Sep;65(9):e27228. [doi: 10.1002/pbc.27228] [Medline: 29797644]
71. One Sickle Cell Disease Voice. URL: https://www.onescdvoice.com/ [accessed 2020-10-02]
72. Perry EL, Carter PA, Becker HA, Garcia AA, Mackert M, Johnson KE. Health literacy in adolescents with sickle cell disease. J Pediatr Nurs 2017;36:191-196. [doi: 10.1016/j.pedn.2017.05.012] [Medline: 28888502]
73. Yee ME, Meyer EK, Fasano RM, Lane PA, Josephson CD, Brega AG. Health literacy and knowledge of chronic transfusion therapy in adolescents with sickle cell disease and caregivers. Pediatr Blood Cancer 2019 Jul;66(7):e27733. [doi: 10.1002/pbc.27733] [Medline: 30941859]
74. Nandyala AS, Nelson LA, Lagotte AE, Osborn CY. An analysis of whether health literacy and numeracy are associated with diabetes medication adherence. Health Lit Res Pract 2018 Jan;2(1):e15-e20 [FREE Full text] [doi: 10.3928/24748307-20171212-01] [Medline: 30112462]
75. Horne R, Chapman SC, Parham R, Freemantle N, Forbes A, Cooper V. Understanding patients' adherence-related beliefs about medicines prescribed for long-term conditions: a meta-analytic review of the Necessity-Concerns Framework. PLoS One 2013;8(12):e80633 [FREE Full text] [doi: 10.1371/journal.pone.0080633] [Medline: 24312488]
76. Horne R, Weinman J, Hankins M. The beliefs about medicines questionnaire: The development and evaluation of a new method for assessing the cognitive representation of medication. Psychol Health 1999;14(1):1-24. [doi: 10.1080/08870449908407311]
77. Horne R, Weinman J. Patients' beliefs about prescribed medicines and their role in adherence to treatment in chronic physical illness. J Psychosom Res 1999 Dec;47(6):555-567. [doi: 10.1016/s0022-3999(99)00057-4] [Medline: 10661603]
78. Badawy SM, Kuhns LM. Economic evaluation of text-messaging and smartphone-based interventions to improve medication adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2016 Oct 25;4(4):e121 [FREE Full text] [doi: 10.2196/mhealth.6425] [Medline: 27780795]
79. Bane C, Hughes CM, McElnay JC. The impact of depressive symptoms and psychosocial factors on medication adherence in cardiovascular disease. Patient Educ Couns 2006 Feb;60(2):187-193. [doi: 10.1016/j.pec.2005.01.003] [Medline: 16253468]
80. Gauchet A, Tarquinio C, Fischer G. Psychosocial predictors of medication adherence among persons living with HIV. Int J Behav Med 2007;14(3):141-150. [doi: 10.1007/bf03000185] [Medline: 18062057]
81. Conn KM, Halterman JS, Fisher SG, Yoos HL, Chin NP, Szilagyi PG. Parental beliefs about medications and medication adherence among urban children with asthma. Ambul Pediatr 2005;5(5):306-310. [doi: 10.1367/A05-004R1.1] [Medline: 16167856]
82. Conn KM, Halterman JS, Lynch K, Cabana MD. The impact of parents' medication beliefs on asthma management. Pediatrics 2007 Sep;120(3):e521-e526 [FREE Full text] [doi: 10.1542/peds.2006-3023] [Medline: 17766496]
83. Bucks RS, Hawkins K, Skinner TC, Horn S, Seddon P, Horne R. Adherence to treatment in adolescents with cystic fibrosis: the role of illness perceptions and treatment beliefs. J Pediatr Psychol 2009 Sep;34(8):893-902. [doi: 10.1093/jpepsy/jsn135] [Medline: 19196850]
84. Gabr WM, Shams ME. Adherence to medication among outpatient adolescents with epilepsy. Saudi Pharm J 2015 Jan;23(1):33-40 [FREE Full text] [doi: 10.1016/j.jsps.2014.05.003] [Medline: 25685041]
85. Goodfellow NA, Hawwa AF, Reid AJ, Horne R, Shields MD, McElnay JC. Adherence to treatment in children and adolescents with cystic fibrosis: a cross-sectional, multi-method study investigating the influence of beliefs about treatment and parental depressive symptoms. BMC Pulm Med 2015 Apr 26;15:43 [FREE Full text] [doi: 10.1186/s12890-015-0038-7] [Medline: 25927329]
86. Klok T, Kaptein AA, Duiverman EJ, Brand PL. High inhaled corticosteroids adherence in childhood asthma: the role of medication beliefs. Eur Respir J 2012 Nov;40(5):1149-1155 [FREE Full text] [doi: 10.1183/09031936.00191511] [Medline: 22362847]
87. World Health Organization. World Health Organization. 2003. Adherence to Long-term Therapies: Evidence for Action URL: http://apps.who.int/iris/bitstream/10665/42682/1/9241545992.pdf [accessed 2019-06-02]
88. Jorgensen DR, Metti A, Butters MA, Mettenburg JM, Rosano C, Novelli EM. Disease severity and slower psychomotor speed in adults with sickle cell disease. Blood Adv 2017 Sep 26;1(21):1790-1795 [FREE Full text] [doi: 10.1182/bloodadvances.2017008219] [Medline: 29296825]
89. Smith KE, Schatz J. Working memory in children with neurocognitive effects from sickle cell disease: contributions of the central executive and processing speed. Dev Neuropsychol 2016;41(4):231-244 [FREE Full text] [doi: 10.1080/87565641.2016.1238474] [Medline: 27759435]
90. Vichinsky EP, Neumayr LD, Gold JI, Weiner MW, Rule RR, Truran D, Neuropsychological Dysfunction and Neuroimaging Adult Sickle Cell Anemia Study Group. Neuropsychological dysfunction and neuroimaging abnormalities in neurologically intact adults with sickle cell anemia. J Am Med Assoc 2010 May 12;303(18):1823-1831 [FREE Full text] [doi: 10.1001/jama.2010.562] [Medline: 20460621]

91. Krauskopf K, Federman AD, Kale MS, Sigel KM, Martynenko M, O'Conor R, et al. Chronic obstructive pulmonary disease illness and medication beliefs are associated with medication adherence. Chronic Obstr Pulm Dis 2015 Apr;12(2):151-164 [FREE Full text] [doi: 10.3109/15412555.2014.922067] [Medline: 24960306]
92. Mayberry LS, Mulvaney SA, Johnson KB, Osborn CY. The MEssaging for diabetes intervention reduced barriers to medication adherence among low-income, diverse adults with type 2. J Diabetes Sci Technol 2017 Jan;11(1):92-99 [FREE Full text] [doi: 10.1177/1932296816668374] [Medline: 27595710]
93. Anie KA. Psychological complications in sickle cell disease. Br J Haematol 2005 Jun;129(6):723-729. [doi: 10.1111/j.1365-2141.2005.05500.x] [Medline: 15952997]
94. Compas BE, Jaser SS, Reeslund K, Patel N, Yarboi J. Neurocognitive deficits in children with chronic health conditions. Am Psychol 2017;72(4):326-338. [doi: 10.1037/amp0000042] [Medline: 28481580]
95. Prussien KV, Jordan LC, DeBaun MR, Compas BE. Cognitive function in sickle cell disease across domains, cerebral infarct status, and the lifespan: a meta-analysis. J Pediatr Psychol 2019 Sep 1;44(8):948-958. [doi: 10.1093/jpepsy/jsz031] [Medline: 31050352]
96. Lenhart A. Pew Internet Research. 2015 Apr 15. Teens, Social Media & Technology Overview 2015 URL: http://www.pewinternet.org/2015/04/09/teens-social-media-technology-2015/ [accessed 2019-05-18]
97. Crosby LE, Barach I, McGrady ME, Kalinyak KA, Eastin AR, Mitchell MJ. Integrating interactive web-based technology to assess adherence and clinical outcomes in pediatric sickle cell disease. Anemia 2012;2012:492428 [FREE Full text] [doi: 10.1155/2012/492428] [Medline: 22701785]
98. Curtis K, Lebedev A, Aguirre E, Lobitz S. A medication adherence app for children with sickle cell disease: qualitative study. JMIR Mhealth Uhealth 2019 Jun 18;7(6):e8130 [FREE Full text] [doi: 10.2196/mhealth.8130] [Medline: 31215518]
99. McGrady ME, Ryan JL, Gutiérrez-Colina AM, Fredericks EM, Towner EK, Pai AL. The impact of effective paediatric adherence promotion interventions: systematic review and meta-analysis. Child Care Health Dev 2015 Nov;41(6):789-802. [doi: 10.1111/cch.12271] [Medline: 26255643]
100. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RC, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 4;157(11):785-795. [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
101. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mHealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014 Oct 24;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
102. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 8;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
103. Stinson JN, Lalloo C, Harris L, Isaac L, Campbell F, Brown S, et al. iCanCope with Pain: User-centred design of a web- and mobile-based self-management program for youth with chronic pain based on identified health care needs. Pain Res Manag 2014;19(5):257-265 [FREE Full text] [doi: 10.1155/2014/935278] [Medline: 25000507]
104. Muessig KE, LeGrand S, Horvath KJ, Bauermeister JA, Hightow-Weidman LB. Recent mobile health interventions to support medication adherence among HIV-positive MSM. Curr Opin HIV AIDS 2017 Sep;12(5):432-441 [FREE Full text] [doi: 10.1097/COH.0000000000000401] [Medline: 28639990]
105. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mHealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019 Apr 6;16(7):pii: E1226 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
106. Bravo C, O'Donoghue C, Kaplan C, Luce J, Ozanne E. Can mHealth improve risk assessment in underserved populations? Acceptability of a breast health questionnaire app in ethnically diverse, older, low-income women. J Health Dispar Res Pract 2014;7(4):pii: 6 [FREE Full text] [Medline: 25705576]
107. Swendeman D, Comulada WS, Koussa M, Worthman CM, Estrin D, Rotheram-Borus MJ, et al. Longitudinal validity and reliability of brief smartphone self-monitoring of diet, stress, and physical activity in a diverse sample of mothers. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e176 [FREE Full text] [doi: 10.2196/mhealth.9378] [Medline: 30249576]
108. Poushter J, Bishop C, Chwe H. Pew Research Center. 2018 Jun 19. Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones URL: http://www.pewglobal.org/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/ [accessed 2019-01-25]
109. Perrin A. Pew Research Center. 2019 May 31. Digital Gap Between Rural and Nonrural America Persists URL: http://www.pewresearch.org/fact-tank/2017/05/19/digitalgap-between-rural-and-nonrural-america-persists/ [accessed 2019-09-26]
110. 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]
111. Campbell BR, Ingersoll KS, Flickinger TE, Dillingham R. Bridging the digital health divide: toward equitable global access to mobile health interventions for people living with HIV. Expert Rev Anti Infect Ther 2019 Mar;17(3):141-144. [doi: 10.1080/14787210.2019.1578649] [Medline: 30721103]

112. Hankins J, Shah N, DiMartino L, Brambrilla D, Fernandez M, Gibson R, Sickle Cell Disease Implementation Consortium. Integration of mHealth into Sickle Cell Disease Care to Increase Hydroxyurea Utilization (mESH) - An Efficacy and Implementation Study. JMIR research protocols 2019:- (forthcoming) [FREE Full text] [doi: 10.2196/preprints.16319]
113. Agency 39, LLC. URL: https://agency39a.com/ [accessed 2020-02-18]

## Abbreviations

**AHRQ:** Agency for Healthcare Research and Quality
**FDA:** Food and Drug Administration
**HBM:** Health Belief Model
**mHealth:** mobile health
**NHLBI:** National Heart Lung and Blood Institute
**NIH:** National Institutes of Health
**NVS:** Newest Vital Sign
**SCCRIP:** Sickle Cell Clinical Research and Intervention Program
**SCD:** sickle cell disease
**SCDIC:** sickle cell disease implementation consortium

*Edited by G Eysenbach; submitted 09.08.19; peer-reviewed by L Crosby, C Jacob; comments to author 17.09.19; revised version received 31.10.19; accepted 26.01.20; published 08.05.20.*

*Please cite as:*
*Alberts NM, Badawy SM, Hodges J, Estepp JH, Nwosu C, Khan H, Smeltzer MP, Homayouni R, Norell S, Klesges L, Porter JS, Hankins JS*
*Development of the InCharge Health Mobile App to Improve Adherence to Hydroxyurea in Patients With Sickle Cell Disease: User-Centered Design Approach*
*JMIR Mhealth Uhealth 2020;8(5):e14884*
*URL: https://mhealth.jmir.org/2020/5/e14884*
*doi:10.2196/14884*
*PMID:32383683*

Original Paper

# Health System Stakeholders' Perspective on the Role of Mobile Health and Its Adoption in the Swiss Health System: Qualitative Study

Myriam Lingg[1*], DPhil; Verena Lütschg[2*], DPhil

[1]SOS Beobachter Foundation, Zurich, Switzerland

[2]About Tomorrow Consulting, Zurich, Switzerland

[*]all authors contributed equally

**Corresponding Author:**
Myriam Lingg, DPhil
SOS Beobachter Foundation
Flurstrasse 55
Zurich, 8048
Switzerland
Phone: 41 058 269 21 21
Email: myriam.lingg@ringier.ch

## Abstract

**Background:** Digital health solutions have great potential to change the way health care is delivered, including better clinical outcomes and improved processes and access to health services. However, the adoption of mobile health (mHealth) solutions for patient monitoring has been rather slow in Switzerland. The reasons are complex, and a better understanding is needed to leverage the full potential of mHealth.

**Objective:** This study aimed to deepen the understanding of the potential relevance and influence of mHealth for the health system and health care provision, and factors influencing its adoption. The findings will be used to provide an outlook on feasible recommendations for action.

**Methods:** We conducted a qualitative survey using a maximum variation sample of a heterogeneous group of stakeholders (N=50) in the Swiss health care system with a profound knowledge of digital health and medical devices. A semistructured interview guide including open- and closed-ended questions was used to address questions around mHealth relevance and its influence on the health system, the relevance of selected determinants for mHealth adoption, and important influencing factors. A content analysis method was applied.

**Results:** Overall, respondents thought that mHealth would have a beneficial impact on the Swiss health system but that its adoption would evolve slowly. We derived 23 key opportunities regarding patient and patient pathway, treatment of disease, and diseases and health conditions. High consistency in answers among respondents was observed for *treatment of disease*. Stakeholders' attitudes toward mHealth adoption along the relevance of 23 preselected determinants were relatively consistent. However, we obtained diverging attitudes regarding the influence of *trends*, *enablers*, and *restraints* in Switzerland and translated them into 26 key themes influencing mHealth adoption. Relevant trends comprise *changing needs and expectations of patients*, *a rising need for efficient health care delivery*, *growing interest in improved outpatient care*, and *emerging technologies and progressing digitization*. Important enablers include *growing demand for new financing schemes and incentive concepts*, *rising demand for comprehensive information on and stronger body of evidence for mHealth use cases*, and *increasing need for easy to use alternate care approaches*. Challenging restraints are *rigidness of thinking and siloed actions of health system actors*, *complexity of changing the existing regulations and structures*, *little understanding of mHealth use and the role of clinicians*, and *risk of further polarization of the population*.

**Conclusions:** This study provides a comprehensive look at mHealth in the Swiss health system. It becomes apparent that strong governance is inevitable to foster a sustainable data strategy and to reconcile the different interests of stakeholders. The use of mHealth will add value but will not necessarily reduce the burden on the system caused by emerging societal needs and changing disease prevalence.

***(JMIR Mhealth Uhealth 2020;8(5):e17315)*** doi:10.2196/17315

**KEYWORDS**

mobile health; mHealth; eHealth; telehealth; telemedicine; digitization; electronic health record; technology

# Introduction

## Background

Digital health solutions have great potential to change the way health care is delivered. This includes better clinical outcomes as well as improved processes and access to health services. As shown in literature, the promise of deploying digital solutions adds value for the patient, provider, and payer [1-5]. Use cases range from remote patient coaching, monitoring, diagnostics, prognosis, and adherence management (patient compliance) to processes that take place between health care providers.

Mobile health (mHealth), a fast-growing field in digital health, is a wireless mobile health app or device that can be used to support different phases of the patient journey [5]. It refers to the collection methods of personal health data (eg, by sensor technology) and their translation into comprehensive information (eg, artificial intelligence-enabled data analysis). The versatile opportunities provide doctors and patients with new insights regarding the patient´s real-time health status or progress of disease. This allows for immediate action and more personalized recommendations [4,5].

The digitization strategy of the Swiss health policy fosters digitization in the health system, comprising important issues which aim to advance, for example, the information technology (IT) infrastructure and adoption of digital solutions such as mHealth [6,7]. However, the adoption of digital solutions and services, according to recent studies, is progressing slowly [8,9]. Today, agreed aims and actions at all governance levels still show neither substantial impact on the hospital landscape nor the clinical practice in primary care [10-12]. However, during health care expert panels and conferences in Switzerland, it is increasingly pointed out where framework conditions have advanced and may allow next digitization steps.

In many countries, mHealth technology is a fast-developing field [13], but good practices in health care to promote its adoption are scarce [14] and the different expectations and needs of multiple stakeholders involved are rarely sufficiently aligned [15]. The adoption depends on an interplay of a complex set of either enabling or hindering factors such as trust of professional end users, administrators, and patients in digital health solutions [16]. Necessary adjustments in the different health system levels along legal, regulatory, technological, and operational dimensions also fall into these factors [1,7,8,15].

Researchers and experts have described relevant themes concerning mHealth adoption in the Swiss health system, thus contributing to a better understanding. These themes include (1) essential fields of action along the different governance levels as mentioned in literature and summarized in Table 1 [5,6,14,17,18], (2) the present relevance and usage of solutions supporting digital health by health care providers and patients [19-21], and (3) needs and requirements of clinicians regarding mHealth use [16,22-24]. Less well explored is the potential mHealth adoption from an integrated perspective of multiple stakeholders that provide health care or shape health care provision.

**Table 1.** Essential fields of action along with the different governance levels.

| Area of action | Aim | Important actors |
|---|---|---|
| Legal framework for the use of mobile health solutions | Regulation of liability risks, a demarcation between a lifestyle and medical device app. | Authority (regulatory, policy, and normative); Associations of health care providers |
| Data privacy and safety | Regulate data transfer and permissions to the data, access rights, permissibility of data transfer to third parties, storage location, and liability issues. | Mobile health developer; normative authorities |
| Evidence of mobile health solutions | Creating trust in mobile health solutions by certification and proof of evidence of mobile health solutions. | Associations of health care providers; mobile health developers |
| Reimbursement for the use of mobile health | Services associated with the use of mobile health must be appropriately included in the tariff and reimbursement catalog. Similarly, virtual consultations should be billable via mobile health apps or online platforms. | Regulatory authority, associations of health care providers |
| Interoperability of mobile health apps | Implementing the use of mandatory standards for the interoperability of mobile health solutions and devices as an important prerequisite for realizing the potential of mobile health. | Authority (normative and policy); Experts in medical informatics and IT[a]; health care providers; mobile health developers |
| Enabling potential mobile health users | Stepwise introduction of digital tools for health care provision contexts and training in the use of digital health solutions. | All stakeholder groups |

[a]IT: information technology.

## Objective

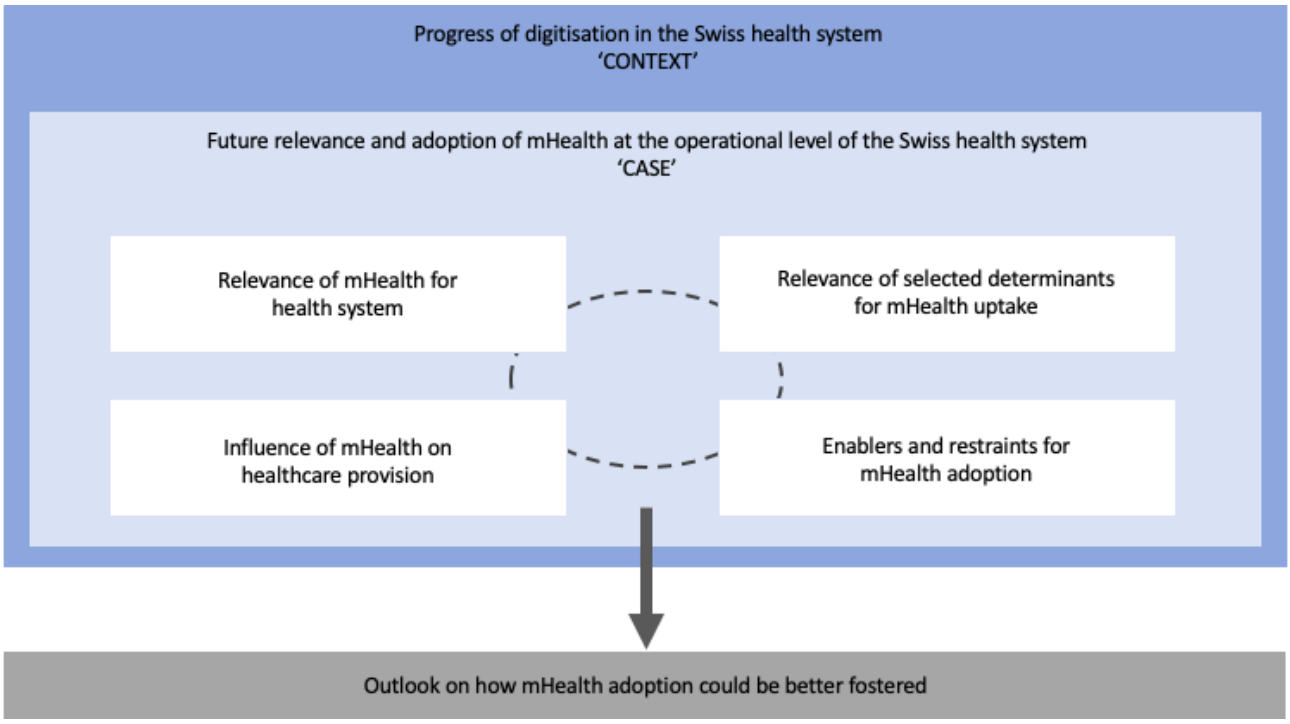
This study aimed to deepen the understanding of potential mHealth adoption in the Swiss health system. To achieve this aim, we assess and evaluate stakeholder perspectives regarding the potential relevance and influence of mHealth for the health system and health care provision, and factors influencing its adoption. The findings will be used to provide an outlook on feasible recommendations for action.

## Methods

### Study Design

We used an embedded case study methodology to integrate quantitative and qualitative methods into a single research study [25,26] by utilizing multiple sources of stakeholders to broaden and deepen data collection, bring together a wealth of data through triangulation and contribute to the validity of the research. These methods helped gain rich insights for this study that focuses on a multifaceted understanding of the future mHealth adoption in the Swiss health system by approaching the same issue from different angles. We have illustrated the main considerations of this approach in Figure 1 which will guide our data collection and analysis.

**Figure 1.** Embedded case study methodology.




We assessed the attitudes of different health system stakeholders toward future mHealth adoption at the operational level of the Swiss health system (*case*). We did this against the background of the progress of digitization in the Swiss health sector (*context*). Further, we based the study design on 4 embedded subunits (*units of analysis*) to answer the study objectives. This gave us the opportunity for a more differentiated analysis.

### Sampling Technique and Participants' Profiles

We used a maximum variation sample [27,28] concentrating on stakeholders that provide health care or shape health care provision, and we applied 2 selection criteria: (1) recruitment of a heterogeneous sample across stakeholders concerned with digital health topics (clinicians, health care organizations, pharmacy, medical device industry, health care start-ups, health sector associations, experts in medical informatics and IT, digital health-related experts, reimbursement-related actors, and government- and research-related bodies); and (2) stakeholders with the ability to provide rich and in-depth information about digitization, electronic health (eHealth), and mHealth, medical device regulations, and reimbursement. We conducted the study with stakeholders from Switzerland and identified interview participants by first searching listings from websites of authorities and similar bodies, Swiss health care start-ups, health technology suppliers, health care organizations, and pharmacies. We then identified authors of reports on eHealth and mHealth policy regulation, regulation and use in Switzerland, and finally, we asked interviewees to recommend other stakeholders. We contacted prospective interview participants by email or LinkedIn between July and September 2019.

### Sample Size and Data Collection Method

In total, 50 interviews were conducted between July and October 2019 by the principal investigator and a research assistant (Table 2).

Interviewees had a choice of being interviewed in German or English. We used a file naming system and anonymized interviewees by generating a list of archival numbers. We conducted face-to-face (n=38) or phone interviews (n=11), and 1 follow-up phone call was based on written participation (n=1). Interviews averaged 36 min (range 23-59 min). Interviews were not audio-recorded, but detailed notes were taken. Written consent was given by all participants, and monetary or other compensation for participation was not provided.

We used a semistructured interview guide. The selection of questions was guided by the experience of the investigators in health systems, health technologies, and digital health. We used the themes of Table 1 and a literature search as an orientation to define the open-ended questions and to select determinants for the closed-ended questions. The guide was based on 4 sets of questions: (1) what is the potential relevance of mHealth for the Swiss health system, (2) what is the potential influence of mHealth on health care provision, (3) what is the relevance of

selected determinants for mHealth adoption, and (4) what are the influencing factors for mHealth adoption. We selected relevant topics for the operational level where health care services are provided to patients.

Before data collection, a semistructured interview guide (Multimedia Appendix 1) was validated based on 2 interviews with 1 clinician and 1 health technology provider and was additionally critically revised by 1 scientist.

**Table 2.** Composition of participants (N=50).

| Main role of participant within the Swiss health care sector | Value, n (%) |
|---|---|
| Providers of health care services (clinicians, health care organizations, and pharmacies) | 9 (18) |
| Providers of health technologies (medical device industry and health care start-ups) | 9 (18) |
| Health sector associations (innovation-promoting associations and interest groups in the health sector) | 7 (14) |
| Consultancy for health system | 7 (14) |
| Digital health-related experts | 5 (10) |
| Experts in medical informatics and IT[a] | 5 (10) |
| Reimbursement-related actors (insurance and insurance association) | 4 (8) |
| Government- and research-related bodies | 4 (8) |

[a]IT: information technology.

### Data Analysis

For the data analysis of open-ended questions and comments provided during closed-ended questions, we thematically analyzed the transcripts based on a content analysis method [26] and MAXQDA software (version 11, VERBI GmbH) was used to aid data management. To begin, both investigators closely read each transcript (data orientation). The main investigator then deductively coded one-third of the transcripts and inductively coded for new themes (data reduction). Following the coding, both investigators revised the list of themes, improved codes, and clustered them into categories (data display). Thereafter, the main investigator systematically applied coding to all transcripts. The assistant investigator critically reviewed a sample of 21 coded transcripts (final coding). Finally, both investigators drew on important themes (conclusion drawing).

For the data analysis of closed-ended questions, we applied descriptive statistics. We grouped the 5 scale values of answers into 3 groups (*high to very high*, *medium*, and *low to very low*) and calculated proportions per stakeholder group. Descriptive analyses were conducted using Excel software version 16.16.11 (190619). The data were tabulated.

### Ethical Considerations

This study did not fall within the scope of the Human Research Act. Therefore, authorization from the ethics committee was not required (BASEC-Nr. Req-2019-01070).

## Results

### Characteristics of Interviewees

The sample included stakeholders with different roles in the health sector (Table 3). About half of the participants (26/50, 52%) indicated to have more than 1 professional role in the health sector (see question 1 of Multimedia Appendix 1). Many interviewees (36/50, 72%) believed they had high to very high knowledge of mHealth. Few interviewees (4/50, 8%) thought they had moderate knowledge of mHealth but high knowledge of medical devices in general.

**Table 3.** Characteristics of interviewees.

| Main role of participant within the Swiss healthcare sector | Very high to high knowledge of mobile health, n (%) | Average knowledge of mobile health, n (%) | Moderate knowledge of mobile health, n (%) |
|---|---|---|---|
| Providers of health care services (clinicians, health care organizations, and pharmacies) | 7 (78) | 1 (11) | 1 (11) |
| Providers of health technologies (medical device industry and health care start-ups) | 8 (89) | 1 (11) | 0 (0) |
| Associations or similar organizations (health and digitization) | 5 (72) | 1 (14) | 1 (14) |
| Consultancy for health system | 5 (71) | 2 (29) | 0 (0) |
| Experts in digitization (health and nonhealth sectors) | 3 (60) | 2 (40) | 0 (0) |
| Experts in medical informatics and IT[a] | 3 (60) | 2 (40) | 0 (0) |
| Reimbursement-related actors (insurance and insurance association) | 2 (50) | 1 (25) | 1 (25) |
| Government- and research-related bodies | 3 (75) | 1 (25) | 0 (0) |
| Total | 36 (72) | 10 (20) | 4 (8) |

[a]IT: information technology.

## Potential Relevance and Influence of Mobile Health for the Health System and Health Care Provision

Among the different stakeholder groups, many interviewees believed that mHealth would gain a moderate to high importance in general and for selected aspects of health care provision (Multimedia Appendix 2). Respondents of the group *providers of health services* were relatively reserved about the relevance of mHealth in the Swiss health system compared with other survey groups. Overall, interviewees thought that mHealth would be highly influential for *patient monitoring*. They argued that clinical experience with some mobile solutions is already rising and demonstrated added value to the treatment pathway. Further, they believed that mHealth would be very influential for *diagnostics* because it could be used as a supporting tool for medical decision-making. Many interviewees saw only limited potential for mHealth in the field of *prognosis* of diseases. They believed that the maturity level of the current generation of mHealth technologies was still very low.

Overall, respondents deduced that the integration of mHealth into medical processes will add value to the patient journey. Interviewees highlighted several opportunities and emphasized a wide range of areas to illustrate the potential influence of mHealth on health care provision (Multimedia Appendix 3). Content analysis resulted in 23 topics that we grouped into *patient and patient pathway*, *treatment of disease*, and *diseases and health conditions*.

The number of topics different stakeholder groups focused on varied; *providers of health care services* mentioned a relatively wide range of topics whereas *government- and research-related bodies* emphasized fewer areas. Two or less stakeholder groups brought up the topics: *offering a wider spectrum of care and improving access to health services* and *improving screening options before stationary interventions*.

Many respondents thought that mHealth could have an impact in terms of *improving health literacy and empowerment of patients*, *increasing health care efficiency*, *establishing new preventive care approaches*, *enabling continuous monitoring*, *complementing and supporting traditional treatment concepts*, and *aiding decision-making based on supportive analysis and diagnostics*. Topics that were of less interest to individual interviewees but still mentioned by respondents were *enabling early detection of health risks*, *contributing to outpatient care*, *making therapies simpler, better controllable and less error-prone*, *fostering disease management*, *generating and access to real-life data*, *improving understanding of disease progress*, and *controlling effectiveness of therapies more closely and enabling early detection of adverse or suboptimal response to treatments*.

## Factors Influencing Mobile Health Adoption

Multimedia Appendix 4 shows the results of closed-ended questions regarding the potential relevance of specific determinants for future mHealth adoption. Of the 50 respondents, 1 interviewee had no specific opinion on the determinants of questions 2 and 6 of the interview guide and answered them with *no opinion*. Multimedia Appendix 5 provides a selection of comments obtained during this part of the interviews.

Multimedia Appendix 6 illustrates the topics that respondents highlighted regarding future mHealth adoption based on open-ended questions. Content analysis resulted in 26 topics that we grouped into *trends*, *enablers*, and *restraints*. Multimedia Appendix 7 provides a selection of comments obtained during interviews.

## Trends

Interviewees emphasized a total of 11 topics that we grouped into the categories *changing needs and expectations of patients in the health system*, *rising need for efficient health care delivery*, *growing interest in supporting and optimizing outpatient care*, and *emerging technologies and progressing digitization in the health sector*.

### *Changing Needs and Expectations of Patients in the Health System*

Many interviewees concluded that patients are in a phase of upheaval driven by their changing needs and referred to similar changes in the lifestyle area. They considered that mHealth provides patients with a promising opportunity to actively shape this phase. They believed that mHealth adoption would be influenced by patients assuming a more consumer-like mindset, being better informed, and demanding more comprehensive information regarding their health status. In addition, interviewees named the introduction of the electronic patient record and clinicians' attitudes toward mHealth as important steps toward mHealth adoption. They were convinced that by increasing utilization of digital tools in general and in the health system, people would gradually become accustomed to sharing their health data for medical purposes.

### *Rising Demand for Efficient Health Care Delivery*

Many interviewees noted that the rising burden on the health system would drive mHealth acceptance. They believed that health care providers would use mHealth as a strategy to increase process efficiency by integrating it into the patient pathway and responding to changing requirements of the health care environment.

They mentioned that cost savings could be realized through remote follow-up of the patients in both the ambulatory sector by the improved exchange of data with health care providers and the stationary sector by shorter inpatient stays. Yet, no respondent had an opinion about how significant the cost savings may be. Some pointed out that instead of cost savings, new costs may arise. They presumed clinicians would integrate mHealth into their treatment regime in the short term but without adapting traditional workflows. Further, they supposed that patients using mHealth would increase their demand and consumption of health care services in the long term. The driving factors would be new health service opportunities and a growing demand for personalized approaches.

### *Growing Interest in Supporting and Optimizing Outpatient Care*

Some interviewees highlighted the *upcoming new opportunities for patients to engage with the health care system*. In addition, they noted that insurance companies and pharmacies showed first attempts to reinvent or adapt their role and business model in their function as health care providers and in response to the digitization in the health system. Respondents emphasized how these 2 players would focus on specific issues that meet the changing needs of patients by utilizing mHealth solutions as a facilitator (eg, preventive care and follow-up advice). Few interviewees highlighted that the role of insurances was controversially discussed among health system stakeholders in general. On the one hand, they had the opportunity to foster health promotion and prevention (introducing awarding programs). On the other hand, they were subject to legal and societal issues (eg, competences and social scoring debate).

Few respondents accentuated that the rising need for integrated solutions in long-term and elderly care would increase the use of mHealth solutions. They felt that the utilization of partially automated and digitally supported health services would be a reasonable solution to meet the growing demand for service and the lack of coordination among caregivers, as well as to finance the resulting costs.

## Enablers

Interviewees emphasized a total of 6 topics that we grouped into categories *growing demand for new financing schemes and incentive concepts for mHealth*, *rising demand for comprehensive information on and stronger body of evidence for mHealth use cases*, and *increasing need for easy to use alternate care approaches*.

### *Growing Demand for New Financing Schemes and Incentive Concepts for Mobile Health*

The majority of interviewees maintained that mHealth adoption would significantly depend on new financing schemes and incentive concepts for mHealth that are currently not established in the health system. They named a wide range of ideas about how novel financing concepts would enable and foster mHealth adoption at the ambulatory level. For instance, some highlighted the importance of the capitation model or a new form of diagnostic-related group. They claimed that new schemes and concepts could motivate clinicians to adopt digital health solutions with the aim to increase the efficiency of their treatments. Others mentioned that value- or incentive-based systems for professionals, as seen by the latest developments in the insurance sector for outpatient units, would be a step in the right direction.

Few interviewees envisioned a concept beyond monetary incentives that would increase in importance in the long run due to changing needs and expectations of patients and clinicians. Interviewees presumed that patients might pay mHealth-related costs out of pocket in the long-term once they were convinced that it promotes their health or improves serious health conditions. Further, they believed that doctors would place more emphasis on the quality of the patient-doctor relation for specific medical situations than on monetary benefits.

### *Rising Demand for Comprehensive Information on and Stronger Body of Evidence for Mobile Health Use Cases*

Many stakeholders believed that beyond the initial mHealth hype, its adoption is increasingly challenged by a weak body of evidence and thus lack of trust in the promised benefits. Interviewees speculated that potential users would only build up confidence in mHealth when clinicians, in general, are more empowered in the use of digital health tools and understand how it influences their workflow. Some interviewees noted that as long as there was no systematic introduction of mHealth based on clinical studies exploring the value proposition along the treatment pathway, it would not play a significant role in medical services. However, some interviewees also believed that the generation of evidence was challenged by 2 aspects: the rapid development of technology and the lack of recognized methods to assess the clinical value of mHealth use. For instance, few mentioned the need for different evidence standard frameworks for digital health technologies and referred to the United Kingdom. They highlighted the role and influence of

research groups and mHealth developers and how they could contribute to a better body of evidence at a disease level.

Further, some interviewees identified the limited information regarding data accuracy and quality (collection and analysis methods used) and noted that it was an important factor for mHealth adoption. They suggested that a quality label or formal body concerned with the quality of mHealth would be necessary to guarantee the quality of mHealth beyond CE certification.

### *Increasing Need for Easy to Use Alternate Care Approaches*

Few interviewees focused on the political discussion regarding public funds for inpatient and outpatient care. They purported that if decision-makers would transfer more funds to the area of home-based care, novel care approaches would be fostered. In return, public budgets could be relieved because better and digitally supported care could contribute to stabilize multimorbid patients and prevent avoidable emergencies.

## Restraints

Interviewees emphasized 9 topics in total that we grouped into the categories *rigidness of thinking and siloed actions of health system actors*, *complexity of changing the existing regulations and structures*, *little understanding of mHealth use and the role of clinicians*, and *risk of polarization of population regarding mHealth use*.

### *Rigidness of Thinking and Siloed Actions of Health System Actors*

Overall, interviewees mentioned that for successful implementation of mHealth a new way of thinking was needed, but they observed opposite behavior. First, instead of using mHealth as an enabler to contribute to integrative health care approaches, interviewees said that health system actors provided health services in silos. Second, instead of promoting the introduction of digital solutions, interviewees thought that health system actors took opposite measures (eg, difficulties of reimbursement of telemedicine services or the recent cancellation of the poly-medication check provided by pharmacists). Third, instead of developing strategies for new financing schemes, interviewees reported that there was a tendency to force the reimbursement of mHealth into the existing structure. Fourth, instead of fostering the trend to open science (eg, open access to research findings and sharing data) and benefitting from shared data to improve treatments and clinical outcomes, interviewees believed that people were stuck in the data privacy discussion and clinicians were trapped in their habitual management of data.

### *Complexity of Changing the Existing Regulations and Structures*

Many interviewees mentioned that unresolved legal issues (eg, liability issues for service providers) and complex regulations may restrain the use of mHealth, and they also referred to the topics listed in Table 1. They indicated that digitization would require an agile mindset, courage, and mutual support. Examples of countries that are digitally more advanced were given to illustrate this while noting the different political frameworks as a major enabler of digitization (eg, Estonia and Singapore). However, interviewees thought that existing regulations should be adapted instead of forcing innovation into structures that have been established for conventional analog approaches.

Furthermore, some interviewees believed that there are still challenging hurdles regarding interconnectivity and IT infrastructure of outpatient and inpatient units limiting the use of mHealth that cannot be solved easily. Moreover, interviewees noted the present inability of health system actors to handle big data, which would be a key leverage point for advanced data analytics.

### *Improving Understanding of Mobile Health Use and the Role of Clinicians*

Many interviewees focused on the demarcation of the contribution of clinicians and mHealth to health care provision. Overall, they thought that mHealth could contribute to most settings except for acute care.

Many interviewees believed that despite the wide range of opportunities mHealth applications may provide; many would continue to depend on the clinician's support and how the patients engage with mHealth solutions (openness and skills). They noted that patients, especially in serious conditions would request face-to-face contact with their treating clinician and would reject services that may create any distance between them and their doctor. In addition, some interviewees supposed that many mHealth solutions for serious diseases would depend on professional instructions and monitoring. Consequently, patients would depend on their clinician's recommendation and support to use such a tool.

Some interviewees focused on the technological limitations of the present mHealth generation. They were convinced that the level of maturity of mHealth solutions with artificial intelligence-enabled data analysis function was still low. Further, few interviewees feared that counteracting effects regarding people's health could emerge from increased use of mHealth for monitoring purposes. They expected that people would tend to feel less responsible themselves and provoke health issues.

### *Risk of Polarization of Population Regarding Mobile Health Use and Counter Effects*

Many interviewees believed that attitudes of older generations of people toward digital health and diverging digital affinity of people regarding the use of digital solutions in general, would manifest in an uneven mHealth adoption. They believed that a growing fragmentation of health service recipients would be observed in the long-term. For instance, the way people respond to incentive systems that reward healthy behavior and good health conditions. Therefore, healthy people would stay healthier, and people in poorer health would be disadvantaged.

# Discussion

## Principal Findings

The aim of this study is to deepen the understanding of potential mHealth adoption in the Swiss health system. For this, we assessed and evaluated stakeholder perspectives regarding the potential relevance and influence of mHealth for the health

system and health care provision, and factors influencing its adoption. The findings will be used to provide an outlook on feasible recommendations for action.

We sought to supplement existing knowledge on enablers and barriers for mHealth use by providing a more differentiated understanding of mHealth adoption in a Swiss context and from different thematic angles that were based on a multifaceted stakeholder perspective. Overall, we found that fostering mHealth adoption is feasible and that it will likely positively influence the health system performance regarding process efficiency and clinical outcome.

We found that mHealth is perceived as a positive development by the large majority of respondents because it could offer multiple opportunities for health care. The respondents believed that mHealth adoption would gradually take place over a longer period and strongly depend on how the patient and physician handle it. Current findings in literature suggest that some areas will likely see more mHealth usage, for example, when it targets telemedicine and patient monitoring [29]. Other areas may evolve slower due to specific requirements and needs of the clinician and patient settings [22], indications, and type of health care utilization.

Our study findings suggest that people have a relatively diverse definition of the added value of mHealth solutions to health care provision. The center of attention of the study participants, which did not include patients, were rather topics that concern health care providers and cost efficiency than topics that could add value to the patient-doctor interface or science (eg, data science). This finding necessitates closer inspection of the patient perspective.

Our study showed that respondents had a high consistency in answering topics that refer to the policy discussion in Switzerland, for instance, regarding access to patient data. However, when interviewees were asked what else will trigger mHealth adoption in Switzerland, new topics were revealed. Respondents paid high attention to changing conditions emerging from societal, technological, environmental, economic, and political domains. This finding is aligned with the current health policy and expert discussions and also with initiatives such as the *Swiss personalized health network* and *Midata* [30]; High attention is addressed to ensure that the patient has access and power over his health data but at the same time fostering a health data sharing culture where data are not owned by profit organizations or enterprises.

The discussed topics are a matter of multiple actors in the health system. The findings suggest that the understanding of future mHealth adoption can be fostered by taking the following aspects into consideration:

- mHealth contribution to bridge the gap between conventional approaches in health care provision and changing conditions will be pivotal for its successful adoption. The systematic introduction of mHealth, a better body of evidence, and the role of novel incentive and financing schemes will be influential.
- The decisive lever will be how well the mHealth solution can build on or connect to existing habits and systems used in medical practice. This has been demonstrated by the examples of mHealth solutions that are already certified by a notified body and used for medical applications and by the findings regarding the Swiss health system.
- Innovative approaches that would imply major digitization steps for health care providers and patients are important but will be less successful in Switzerland in the near future. The high complexity of the Swiss health care system makes it difficult to change existing regulations and structures and at the moment it does not offer the required flexibility to create the necessary framework conditions for such innovations.
- In the triangle of patients, providers, and payers, mHealth adoption is influenced by the implications of the deeply entrenched roles of these actors in health care and their tendency to execute health care provision in a traditional way. In consequence, an innovative mindset and novel health care approaches and settings cannot develop easily.

## Outlook

The digitization hype has led stakeholders to evaluate legal, regulatory, and technological framework conditions across all health system levels to deepen the understanding on how to exploit the potential of emerging technologies and to promote digitization in health [31]. However, as technology is evolving in 3 dimensions (advanced materials, biotechnologies, and digital technologies), health care is experiencing ever greater difficulties in responding to the rapid development of digital solutions. One reason for this is the complexity of changing existing structures in the Swiss health system. On the other hand, health system actors require time to understand the broad spectrum of opportunities for emerging technologies. They have to develop knowledge by reaching out for professional support, which is often lacking.

Findings from other research and recent developments in the health system indicate that discussions and proposals by stakeholders are seeing gradual developments on how to promote mHealth adoption from different angles. For instance, considering social and organizational factors [24] and adopting a holistic approach for the development of digital health solutions [32]. However, a wide range of issues that remain as challenges to mHealth adoption are evolving to increasingly crucial barriers. For instance, solving policy discussions regarding self-determination of patient information, managing digital communication embedded in complex scenarios and treatment pathways, and addressing implications of the rising number of software-based medical devices [33-36].

Digitization in the Swiss health system will take place stepwise as it is an ongoing process of understanding and integrating emerging technology. It highly depends on how the culture of health care actors and patients evolves regarding the adoption and management of digital health solutions [37]. The 2019 commonwealth fund study highlights, for example, that [38]:

- The proportion of Swiss clinicians (69.7%) who document the medical history electronically is still very low.
- Only 46.6% of the Swiss clinicians consider supporting the use of the electronic patient record system.

- A total of 46.5% of Swiss clinicians exchange clinical data of their patients with clinicians from outside of their office.

As long as the level of digitization in the health sector is low, mHealth adoption will progress slowly. We do not only have to manage the expectations regarding added value but also potential drawbacks of the use of novel digital health solutions when they fail as seen recently with a diabetes monitor [39].

On the basis of the findings of our study, recent research, and policy discussion, we reveal an outlook on how mHealth adoption could be better promoted by approaching the topic from new angles and thus beyond the already identified restraints and defined actions:

- Comprehensive information and strong evidence regarding specific mHealth solutions uncovering relevant potentials and limitations.
- Better communication by interest groups and media about the broad application fields of mHealth solutions to support patients, providers, and payers.
- Active patient lobbying to better represent the needs and expectations of patients.
- Strong governance to establish long-term perspectives for the use of digital health technologies and strategies that give the actors room for actions.
- Open discussion and education to overcome barriers that are rooted in the culture of traditional health care along the triangle of patients, providers, and payers.
- Innovative approaches across stakeholders to break down rigid structures and to empower and enable the use and integration of digital health solutions.
- New approaches of cooperation at the interfaces of the triangle of health care recipients, providers, and insurer that provides added value to all involved stakeholders.

## Limitations

Whereas this study contributes to deepening the understanding of factors influencing mHealth adoption in Switzerland, some limitations have to be acknowledged. We used a qualitative method which does not necessarily guarantee the sample being representative for the population of stakeholders involved in the Swiss health system. This study included a variety of stakeholders in terms of expertise and role within the health system, but not all possible interest groups could be considered. Even though patients' needs and demands are important in the progression of mHealth adoption, they have not been included because their recruitment proved to be very challenging and the focus of this study is on stakeholders that provide health care or shape health care provision (at a system level). Moreover, the sampling was based on a maximum variation strategy and may constitute a selection bias. The interpretation of the findings that not only served to assess enablers and restraints of mHealth adoption but also to define an outlook on how to promote mHealth adoption was a subjective process.

## Conclusions

This study provides an analysis of mHealth adoption in Switzerland from new perspectives. What is becoming increasingly apparent beyond the digital hype, however, is that governance in general and structured data, in particular, are becoming more important. Well-executed health data coordination and exchange are crucial to internalize the added value of new and digitally supported health care environments and ecosystems. The introduction of the Swiss electronic patient record will be an important step forward, but it only provides a formal framework for an advanced playground of health care stakeholders. The adoption of different types of digital health solution is not necessarily disrupting the health system but transforming it to a certain degree. Governance at the different levels of the health system plays a central role in reconciling the different interests of stakeholders and multifaceted impacts that emerge from changing conditions. Digital solutions promise to increase efficiency, contribute to treatment effectiveness, and to improve the mode of communication between patients and health care providers. However, these solutions will not necessarily solve the burden on the system caused by emerging societal needs and changing disease prevalence. Behavioral change of the society and change of habits of stakeholders will also be necessary to internalize the positive effects of digitization. This study provides an outlook on how mHealth adoption could be better promoted by approaching the topic from new angles. Thus, it may contribute to enriching decision-making and actions of policy makers and other stakeholders who have the aim of fostering the adoption of digital solutions into health care.

## Acknowledgments

The authors are particularly grateful to all the people who participated in this study, as without their involvement, this investigation would not have been possible. The authors greatly appreciate the scientific support of Prof Kaspar Wyss (Swiss Tropical and Public Health Institute) who has been involved in revising the manuscript critically for important intellectual content and structure. The authors also appreciate the support of the Swiss startup *evismo AG* with whom they reflected on the findings of their study. The authors declare that they do not have an external funding source. The corresponding author and the coauthors have not received any monetary support to realize this specific research.

## Authors' Contributions

All authors were involved in the outline of the paper. ML has made substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, and drafting the manuscript. VL has been involved in the acquisition of data, analysis and interpretation of data, and drafting the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File , 84 KB - mhealth_v8i5e17315_app1.docx ]

Multimedia Appendix 2
Potential relevance and influence of mobile health.
[DOCX File , 15 KB - mhealth_v8i5e17315_app2.docx ]

Multimedia Appendix 3
Topics illustrating the potential influence of mobile health on health care provision.
[DOCX File , 22 KB - mhealth_v8i5e17315_app3.docx ]

Multimedia Appendix 4
Potential relevance of selected determinants regarding the future mHealth adoption.
[DOCX File , 24 KB - mhealth_v8i5e17315_app4.docx ]

Multimedia Appendix 5
Comments during the collection of attitudes based on closed-ended questions.
[DOCX File , 24 KB - mhealth_v8i5e17315_app5.docx ]

Multimedia Appendix 6
Influencing factors for mobile health adoption.
[DOCX File , 51 KB - mhealth_v8i5e17315_app6.docx ]

Multimedia Appendix 7
Selection of comments on trends, enablers and restraints obtained during interviews.
[DOCX File , 30 KB - mhealth_v8i5e17315_app7.docx ]

## References

1. Böttinger E, zu Putlitz J. Die Zukunft der Medizin: Disruptive Innovationen revolutionieren Medizin und Gesundheit. Berlin: MWV Medizinisch wissenschaftlicher Verlag; 2019.
2. Fischer M, Hafen E, Jelitto J, Kaiserswerth M, Kossmann D, Marchiori C. Big Data im Gesundheitswesen. Swiss academics communications 2015;10(2) [FREE Full text]
3. Sanyal C, Stolee P, Juzwishin D, Husereau D. Economic evaluations of eHealth technologies: A systematic review. PLoS One 2018;13(6):e0198112 [FREE Full text] [doi: 10.1371/journal.pone.0198112] [Medline: 29897921]
4. World Economic Forum. 2017. Value in Healthcare. Laying the Foundation for Health System Transformation URL: http://www3.weforum.org/docs/WEF_Insight_Report_Value_Healthcare_Laying_Foundation.pdf [accessed 2019-09-01]
5. eHealth Suisse. Bern: eHealthSuisse; 2017 Mar 16. mobile Health (mHealth) Empfehlungen I: Ausgangslage und erste Schritte URL: https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/2017/D/170316_mHealth_Empfehlungen_I_d.pdf [accessed 2019-09-01]
6. eHealth Suisse. Bern: eHealthSuisse URL: https://www.e-health-suisse.ch/de/startseite.html [accessed 2019-09-01]
7. Bundesamt für Gesundheit BAG. Bern: Bundesamt für Gesundheit BAG Gesundheit 2020 URL: https://www.bag.admin.ch/bag/de/home/strategie-und-politik/gesundheit-2020.html [accessed 2019-09-01]
8. Thiel R, Deimel L, Schmidtmann D, Piesche K, Hüsing T, Rennoch J. Bertelsmann Stiftung. Berlin: Bertelsmann Stiftung; 2018 Nov 29. Digitalisierungsstrategien im internationalen Vergleich URL: https://tinyurl.com/y9vpzudv [accessed 2019-09-01]
9. Digital Swiss. Bern: Digital.Swiss Digitale Innovation in der Schweiz. Thema Gesundheit URL: https://digital.swiss/de/themen/gesundheit [accessed 2019-09-01]
10. Feldges D. Neue Zürcher Zeitung. Zürich: Neue Zürcher Zeitung; 2019 Mar 28. Bei der Digitalisierung des Gesundheitswesens ist Skandinavien der Schweiz um Längen voraus URL: https://www.nzz.ch/wirtschaft/skandinavien-digitalisierung-im-gesundheitswesen-weit-voraus-ld.1470452 [accessed 2019-09-01]
11. Medinside.: Medinside. Das Portal für die Gesundheitsbranche; 2019 Apr 17. Digitale Innovation im Gesundheitswesen URL: https://www.medinside.ch/de/post/smarter-healthcare-fuer-digitale-innovation-im-gesundheitswesen [accessed 2019-09-01]

12. Zingg T, Sojer R, Röthlisberger F. Digitalisierung in der ambulanten Gesundheitsversorgung. Schweiz Ärzteztg 2019 Jan 30;100(5):113-116 [FREE Full text] [doi: 10.4414/saez.2019.17521]
13. Research 2 Guidance. Berlin: Research2Guidance; 2017. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health URL: https://tinyurl.com/y9vsk8bc [accessed 2019-09-01]
14. European Commission. 2015. Summary Report on the Public Consultation on the Green Paper on Mobile Health URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=8382 [accessed 2019-09-01]
15. van Velthoven MH, Cordon C. Sustainable adoption of digital health innovations: perspectives from a stakeholder workshop. J Med Internet Res 2019 Mar 25;21(3):e11922 [FREE Full text] [doi: 10.2196/11922] [Medline: 30907734]
16. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. J Med Internet Res 2018 Dec 13;20(12):e11254 [FREE Full text] [doi: 10.2196/11254] [Medline: 30545807]
17. Meidert U, Becker H, Scheermesser M, Prieur Y, Hegyi S, Stockinger K, et al. Quantified Self - Schnittstelle zwischen Lifestyle und Medizin. Zurich: Vdf Hochschulverlag Ag; 2018.
18. Endel R, Jäschke T, Thiel C, Wickinghoff D. eHealth Suisse. Bern: eHealthSuisse; 2015 Mar 19. mHealth im Kontext des elektronischen Patientendossiers URL: https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/2015/D/150309_Bericht_FHSG_mHealth_D.pdf [accessed 2019-09-01]
19. Golder L, Jans C. Swiss eHealth Barometer 2019. Bern: gfs.bern; 2019 Feb. Bericht zur Bevölkerungsbefragung URL: https://tinyurl.com/y9lzny8d [accessed 2019-09-01]
20. Lienhard S. Die Digitalisierung in Schweizer Spitälern. Zurich: Hochschule für Wirtschaft Zürich; 2018 Jul 02. Der digitale Reifegrad der Schweizer Kliniken und Spitäler URL: https://www.medinside.ch/de/post/spitaeler-stehen-punkto-digitalisierung-erst-am-anfang [accessed 2019-09-01]
21. Bareuther V. ZHAW digitalcollection. Zurich: Zürcher Hochschule für angewandte Wissenschaften; 2017 Jul 7. Entwicklungen eines spitalbezogenen eHealth Reifegradmodells und Validierung am Beispiel Hirslanden URL: https://digitalcollection.zhaw.ch/bitstream/11475/7704/1/Bareuther_Verena_W.MA.BA.HE.pdf [accessed 2019-09-01]
22. Jacob C, Sanchez-Vazquez A, Ivory C. Clinicians' role in the adoption of an oncology decision support app in Europe and its implications for organizational practices: qualitative case study. JMIR Mhealth Uhealth 2019 May 3;7(5):e13555 [FREE Full text] [doi: 10.2196/13555] [Medline: 31066710]
23. Medinside. Winterthur; 2019 Aug 20. Meineimpfungen wird europaweit als Leuchtturm-Projekt anerkannt URL: https://www.medinside.ch/de/post/meineimpfungen-wird-europaweit-als-leuchtturm-projekt-anerkannt [accessed 2019-09-01]
24. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
25. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11:100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
26. Mayring P. Qualitative inhaltsanalyse. In: Mey G, Mruck K, editors. Handbuch Qualitative Forschung in der Psychologie. Wiesbaden: VS Verlag für Sozialwissenschaften; 2010:601-613.
27. Hardon A, Hodgkin C, Fresle D. How to investigate the Use of Medicines by Consumers. Geneva: World Health Organization Maximum variation technique URL: http://apps.who.int/medicinedocs/en/d/Js6169e/7.3.html [accessed 2019-09-01]
28. Yin R. Journal of Research in Human Resource Management.: Sage Publications, Ltd; 2009. Case Study Research: Design and Methods URL: https://www.jstor.org/stable/23279888?seq=1 [accessed 2019-09-01]
29. Paglialonga A, Patel A, Pinto E. The healthcare system perspective in mHealth. In: Andreoni G, Perego P, Frumento E, editors. M_Health Current and Future Applications. Berlin: Springer; 2019:127-142.
30. Hafen E. ETH Zürich. Zurich: ETHZ; 2019 Mar 14. Die europäische Datenmisere und die personalisierte Medizin URL: https://ethz.ch/de/news-und-veranstaltungen/eth-news/news/2019/03/blog-hafen-europaeische-datenmisere.html [accessed 2020-02-05]
31. [Swiss Confederation].: Bundesamt für Gesundheit BAG [The Federal Council's health policy strategy 2020–2030] URL: https://www.newsd.admin.ch/newsd/message/attachments/59488.pdf [accessed 2019-12-10]
32. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 5;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
33. Mäder L. Neue Zürcher Zeitung. Zurich: Neue Zürcher Zeitung; 2019 Sep 6. Politiker streiten um Datenschutz – die Schweiz könnte mit der EU Probleme bekommen URL: https://tinyurl.com/ybyghc4w [accessed 2019-12-04]
34. Der Bundesrat. Bern: Bundesamt für Gesundheit BAG; 2019 Jan 28. Datenschutztag 2019: Drei Schwerpunkte für Bund und Kantone URL: https://www.admin.ch/gov/de/start/dokumentation/medienmitteilungen.msg-id-73775.html [accessed 2019-12-04]
35. Bürkle T, Lehmann M, Denecke K. Healthcare of the Future: Bridging the Information Gap. Amsterdam: IOS Press; 2019.
36. Darms M, Haßfeld S, Fedtke S. Medizintechnik und medizinische Geräte als potenzielle Schwachstelle. In: Darms M, Haßfeld S, Fedtke S, editors. IT-Sicherheit und Datenschutz im Gesundheitswesen. Berlin: Springer; 2019:109-128.

37. Angerer A, Russ C, Ultsch A. ZHAW digitalcollection. Winterthur: ZHAW School of Management and Law; 2019 Sep. Digital Health - Revolution or Evolution? Strategische Optionen im Gesundheitswesen URL: https://digitalcollection.zhaw.ch/bitstream/11475/18267/1/2019_Digital%20Health%20Report.pdf [accessed 2019-12-04]
38. eHealth Suisse. Bern: Bundesamt für Gesundheit BAG; 2019 Dec 03. Studie des Commonwealth-Fund 2019: Resultate bei eHealth-Instrumenten URL: https://www.e-health-suisse.ch/fileadmin/user_upload/Dokumente/D/factsheet-commonwealth-studie-ehealth-schweiz.pdf [accessed 2020-02-05]
39. O'Connor A. The New York Times. New York: The New York Times; 2019 Dec 2. In Weekend Outage, Diabetes Monitors Fail to Send Crucial Alerts URL: https://www-nytimes-com.cdn.ampproject.org/c/s/www.nytimes.com/2019/12/02/well/live/Dexcom-G6-diabetes-monitor-outage.amp.html [accessed 2019-12-11]

## Abbreviations

**eHealth:** electronic health
**mHealth:** mobile health
**IT:** information technology

*Edited by G Eysenbach; submitted 17.12.19; peer-reviewed by C Cordon, C Jacob, A Adjekum, I Contreras, A Robinson, T Muto; comments to author 02.02.20; revised version received 20.02.20; accepted 22.02.20; published 11.05.20.*

*Please cite as:*
*Lingg M, Lütschg V*
*Health System Stakeholders' Perspective on the Role of Mobile Health and Its Adoption in the Swiss Health System: Qualitative Study*
*JMIR Mhealth Uhealth 2020;8(5):e17315*
*URL: https://mhealth.jmir.org/2020/5/e17315*
*doi:10.2196/17315*
*PMID:32391802*

Original Paper

# Remote Patient Monitoring Technologies for Predicting Chronic Obstructive Pulmonary Disease Exacerbations: Review and Comparison

Kathleen G Fan[1], BSc; Jess Mandel[2], MD; Parag Agnihotri[3], MD; Ming Tai-Seale[4], MPH, PhD

[1]New York Medical College, Valhalla, CA, United States

[2]Division of Pulmonary, Critical Care, and Sleep Medicine, School of Medicine, University of California San Diego, La Jolla, CA, United States

[3]University of California San Diego Health Population Health Services Organization, University of California San Diego, La Jolla, CA, United States

[4]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, United States

**Corresponding Author:**
Ming Tai-Seale, MPH, PhD
Department of Family Medicine and Public Health
University of California San Diego
9500 Gilman Drive
La Jolla, CA
United States
Phone: 1 650 815 5485
Email: mtaiseale@health.ucsd.edu

## Abstract

**Background:** Chronic obstructive pulmonary disease (COPD) is the third leading cause of death by disease worldwide and has a 30-day readmission rate of 22.6%. In 2015, COPD was added to the Medicare Hospital Readmission Reductions Program.

**Objective:** The objective of this paper was to survey the current medical technologies for remote patient monitoring (RPM) tools that forecast COPD exacerbations in order to reduce COPD readmissions.

**Methods:** We searched literature and digital health news to find commercially available RPM devices focused on predicting COPD exacerbations. These technologies were reviewed and compared according to four criteria: forecasting ability, cost, ease of use, and appearance. A rating system was developed to facilitate the evaluation process.

**Results:** As of June 2019, a list of handheld and hands-free devices was compiled. We compared features and found substantial variations. Devices that ranked higher on all criteria tended to have a high or unlisted price. Commonly mass-marketed devices like the pulse oximeter and spirometer surprisingly fulfilled the least criteria.

**Conclusions:** The COPD RPM technologies with most technological promise and compatibility with daily living appear to have high or unlisted prices. Consumers and providers need better access to product information to make informed decisions.

***(JMIR Mhealth Uhealth 2020;8(5):e16147)*** doi:10.2196/16147

**KEYWORDS**

COPD; disease exacerbation; remote patient monitoring; mobile health; telehealth; at-home monitoring; remote monitoring system; wearable

## Introduction

Chronic obstructive pulmonary disease (COPD) is the third leading cause of death worldwide and has a 30-day hospital readmission rate of 22.6% [1,2]. Its exacerbations, if not well managed, can reduce patients' quality of life and increase costs of care [3-6]; COPD is a target condition in the Medicare Hospital Readmission Reductions Program.

Early COPD exacerbation recognition has been a focus of inquiries since faster intervention correlates with better outcomes [7-9]. Oxygen saturation, respiratory rate (RR), and heart rate (HR) have been identified as useful biomarkers [10-14]. Measuring oxygen saturation alone, however, may not be sufficient because it naturally fluctuates throughout the day [14,15]. Measuring these biomarkers in the form of remote patient monitoring (RPM) technologies has improved patients'

abilities to self-manage and decreases COPD's economic and clinical burden [9,16,17].

This paper reviews RPM technologies targeting early COPD exacerbation markers. We provide a concise, practical comparison of technologies to assist consumers and providers in making informed decisions about available technologies.

## Methods

We examined available and commercialized devices by searching for COPD RPM technologies in the literature. We searched Medline and ClinicalTrials.gov and digital health news outlets such as MobiHealthNews, Wearable Technologies, and FierceBiotech with keywords "COPD," "COPD exacerbation," "AECOPD," and "chronic obstructive pulmonary disease" in combinations with "remote patient monitoring," "at-home monitoring," "remote monitoring system," "device," "wearable," and "technology" in June 2019. We narrowed the focus to FDA-approved devices or ones slated for approval within a year. We grouped selected technologies into handheld versus hands-free devices and compared them using four additional criteria: forecasting ability, cost, ease of use, and appearance. Forecasting ability is the likelihood the device would signal a developing COPD exacerbation based on indicative biomarkers [10-14]. Costs were compared at retail prices, if listed. Ease of use was assessed based on efforts required from users to operate the device because patients are more likely to consistently use passive and user-friendly devices [18-20]. Appearance was considered because it influences user acceptability. Unattractive or uncomfortable devices result in lower take-up [18-20].

Using a star scheme, we rated each device for all criteria besides cost where actual amount (if available) was considered. One star indicates poor fulfillment, three stars indicates adequate fulfillment, and five stars indicates excellent fulfillment of that criterion (Table 1).

**Table 1.** Comparison of selected chronic obstructive pulmonary disease handheld and hands-free remote monitors.

| Device | Forecasting ability | Cost | Ease of use | Appearance |
|---|---|---|---|---|
| **Handheld** | | | | |
| Spirometer | ★ | $99-$2500 | ★ | ★★★ |
| Pulse oximeter | ★ | $15-$599 | ★ | ★ |
| Propeller Health sensor | ★ | unlisted | ★★★ | ★★★ |
| Cohero Health kit | ★★★ | $49/mo | ★★★ | ★★★ |
| **Hands-free** | | | | |
| Spry Health Loop System | ★★★★★ | unlisted | ★★★★★ | ★★★★★ |
| Omron HeartGuide | ★★★ | $499 | ★★★★★ | ★★★★★ |
| Spire Health Tag | ★★★★★ | $49 | ★★★ | ★★★ |
| Cosinuss One | ★ | $146.50 | ★★★ | ★★★ |
| Current Health Armband | ★★★★★ | $199 + $40/mo | ★★★ | ★ |
| Adamm RSM | ★★★★★ | unlisted | ★★★★★ | ★★★★★ |

## Results

### Handheld Monitors

Spirometers are the gold standard pulmonary function test for COPD diagnosis. With the rise of over-the-counter tabletop to handheld spirometers, at-home spirometry has become common for daily monitoring of the amount and/or speed of air that can be inhaled and exhaled. Spirometers range in cost from a few hundred to several thousand dollars [21,22].

Pulse oximeters are lightweight devices measuring oxygen saturation that have been mass-marketed for at-home care. Pulse oximeters can cost from $15 to $599 [23,24].

The Propeller Health sensor is an electronic inhaler attachment tracking medication use and potential exacerbations (≥10 inhaler puffs within a 24-hour period or greater usage over a 48-hour period) [25,26]. A study showed 17 of a 39-patient cohort adhered to using the device over a 2-year study [27]. The price of Propeller's asthma sensor is negotiated with health care delivery organizations or payers. It can be free to patients through sponsored health plans but is otherwise around $300 [28]. The price of its COPD sensor is not listed.

The Cohero Health kit includes a medical-grade handheld spirometer, medication-tracking sensor, and web app that centralizes the data [29,30]. The kit comes as a subscription service for $49 per month [31].

### Hands-Free Monitors

Spry Health's Loop System is a wristband monitor. It tracks oxygen saturation, HR, RR, and blood pressure, alerting on significant changes in the wearer's physiological data [32]. Its price is unlisted.

Omron's HeartGuide is a watch-like monitor measuring HR, blood pressure, physical activity, and sleep quality. These features are transmitted to a mobile app where patients can track progress and access health coaching [33]. Although the HeartGuide directly targets patients with heart conditions, its ability to track HR, physical activity, and sleep quality can help detect signs of COPD exacerbation [34,35]. The HeartGuide costs $499.

The Spire Health Tag is a disposable adhesive sensor that attaches to clothing. Each tag tracks RR, HR, breathing pattern, sleep quality, and physical activity, which are logged into a smartphone app [36]. The tags have a 1-year battery life and are washer and dryer safe. Each tag costs $49. Consumers can opt for a Spire membership, which at $10 per month provides free replacements and additional tags for $25 each [37].

The Cosinuss One is a monitor placed in the ear that connects to a smartphone app to measure HR, HR variability, and body temperature. Cosinuss is currently developing OxMotion, an add-on specific to COPD patients tracking RR and oxygen saturation levels. We primarily discuss Cosinuss One because the OxMotion is not yet available. The Cosinuss One costs $146.50 [38].

Current Health's Remotely Monitor system is an armband monitor measuring HR, RR, skin temperature, oxygen saturation, and movement. Data are transmitted to a cloud platform and can be integrated into the patient's electronic health record. At the most basic membership, the Current Armband is $199 upfront and $40 per month for continuous service [39].

The Adamm RSM from Health Care Originals is an adhesive device for the upper torso that monitors cough rate, respiratory patterns, HR, and temperature. Data go to a mobile app and web portal viewable by patients and their physicians [40]. Its price is unlisted.

## Discussion

### Principal Findings

Selected COPD RPM devices were assessed based on forecasting ability, cost, ease of use, and appearance. Spry Health Loop System and Adamm RSM ranked highest across most dimensions aside from cost. The pulse oximeter fulfilled the least criteria.

The Loop System, Current Health Armband, Spire Health Tag, and Adamm RSM ranked highest in forecasting ability since they monitor the most indicative biomarkers. Although Cosinuss One tracks multiple biomarkers, it has shown inaccurate measurements in external studies and thus may not have high forecasting ability [41-43]. Adamm RSM and Spire Health Tag track respiratory patterns, which may be even more accurate [12].

Spirometers, although a well-established COPD diagnostic method, ranked low in forecasting ability. Spirometry, when conducted in outpatient settings or unaided by a health care professional, can often yield inaccurate results due to technical factors [7,44]. Cohero Health's mobile spirometer is International Organization for Standardization (ISO) 9001 and 13485 certified, giving it more accuracy than other mass-marketed spirometers [30]. Although the Propeller sensor is easy to use and has been used to predict incoming exacerbations when it detects increased use of the inhaler, it does not provide data on pulmonary function considered important to predict an incoming exacerbation [27]. Both Cohero and Propeller require patients to use the inhaler, which could face challenges in adherence [18,26].

The Loop System and HeartGuide ranked highest in ease of use. Both devices are wristbands that can be worn without effortful engagement from users to obtain data, which increases usability and adherence [45]. Adamm RSM similarly can be easily hidden under shirts. Other devices with three stars required a little more adjusting. Current Health Armband and pulse oximeter are easy to wear but may shift throughout the day due to their locations on the arm and finger, respectively.

Although relatively easy to use, the Current Health Armband ranked lowest in appearance due to its bulky design and high-profile arm placement. Pulse oximeters also ranked low due to high visibility. While patients can choose to spot check rather than continuously wear the device, this would require some reminder, therefore decreasing its ease of use.

Other devices were relatively low profile. The Loop and HeartGuide are modeled after watches while the handheld spirometer and Propeller inhaler can be carried in a pocket or bag. The Spire Health Tag is out of view once attached to clothing. Adherence becomes challenged if consumers change or discard clothing with the Health Tag attached, however. The Cosinuss One is a low-profile earpiece but may be disadvantageous because it may impede hearing.

Cost-wise, the Spire Health Tag seemingly ranks lowest. But because the sensor may only be attached once, users must decide which clothing they wear most frequently and wear those items every day. Over-the-counter options such as spirometers and pulse oximeters, although inexpensive compared to other RPM devices, are often inaccurate in measurement and have less robust premarket testing [7,46].

### Limitations

This review is not exhaustive. It provides a review of devices currently on the market and readily searchable online. Some technologies were not selected because they were similar in function but not yet on the market or lacked product information. Future efforts are necessary to update this review, given digital health technology is continually improving and evidence for efficacy of RPM in reducing COPD exacerbations is still developing [47]. Another limitation concerns the star rating system, which is based on qualitative assessments rather than quantitative metrics [47]. Last, technology is only one aspect of COPD management strategy. More guideline-concordant treatment and better patient engagement are needed, but they are beyond the scope of this brief report [48].

### Conclusion

Patients can better manage their COPD with the aid of RPM technology that can be easily adopted into their daily routine. The most promising devices are either expensive or without available cost information. Consumers and health care organizations can benefit from more publicly accessible information on COPD RPM products and their comparative effectiveness and costs.

## Conflicts of Interest

None declared.

## References


1. World Health Organization. 2018 May 24. The top 10 causes of death URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2020-04-13]
2. Shah T, Press VG, Huisingh-Scheetz M, White SR. COPD readmissions: addressing COPD in the era of value-based health care. Chest 2016 Oct;150(4):916-926 [FREE Full text] [doi: 10.1016/j.chest.2016.05.002] [Medline: 27167208]
3. McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. Circulation 2015 May 19;131(20):1796-1803 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.010270] [Medline: 25986448]
4. Blanchette CM, Gross NJ, Altman P. Rising costs of COPD and the potential for maintenance therapy to slow the trend. Am Health Drug Benefits 2014 Apr;7(2):98-106 [FREE Full text] [Medline: 24991394]
5. Goto T, Faridi MK, Gibo K, Toh S, Hanania NA, Camargo CA, et al. Trends in 30-day readmission rates after COPD hospitalization, 2006-2012. Respir Med 2017 Sep;130:92-97 [FREE Full text] [doi: 10.1016/j.rmed.2017.07.058] [Medline: 29206640]
6. Wilkinson TMA, Donaldson GC, Hurst JR, Seemungal TAR, Wedzicha JA. Early therapy improves outcomes of exacerbations of chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2004 Jun 15;169(12):1298-1303. [doi: 10.1164/rccm.200310-1443OC] [Medline: 14990395]
7. Liistro G, Vanwelde C, Vincken W, Vandevoorde J, Verleden G, Buffels J, COPD Advisory Board. Technical and functional assessment of 10 office spirometers: a multicenter comparative study. Chest 2006 Sep;130(3):657-665. [doi: 10.1378/chest.130.3.657] [Medline: 16963659]
8. Borel J, Pelletier J, Taleux N, Briault A, Arnol N, Pison C, et al. Parameters recorded by software of non-invasive ventilators predict COPD exacerbation: a proof-of-concept study. Thorax 2015 Mar;70(3):284-285. [doi: 10.1136/thoraxjnl-2014-206569] [Medline: 25582449]
9. Hurst JR, Skolnik N, Hansen GJ, Anzueto A, Donaldson GC, Dransfield MT, et al. Understanding the impact of chronic obstructive pulmonary disease exacerbations on patient health and quality of life. Eur J Intern Med 2020 Jan 16 [FREE Full text] [doi: 10.1016/j.ejim.2019.12.014] [Medline: 31954592]
10. Shah SA, Velardo C, Farmer A, Tarassenko L. Exacerbations in chronic obstructive pulmonary disease: identification and prediction using a digital health system. J Med Internet Res 2017 Mar 07;19(3):e69 [FREE Full text] [doi: 10.2196/jmir.7207] [Medline: 28270380]
11. Yañez AM, Guerrero D, Pérez DAR, Garcia-Rio F, Alvarez-Sala JL, Calle-Rubio M, et al. Monitoring breathing rate at home allows early identification of COPD exacerbations. Chest 2012 Dec;142(6):1524-1529. [doi: 10.1378/chest.11-2728] [Medline: 22797131]
12. Blouet S, Sutter J, Fresnel E, Kerfourn A, Cuvelier A, Patout M. Prediction of severe acute exacerbation using changes in breathing pattern of COPD patients on home noninvasive ventilation. Int J Chron Obstruct Pulmon Dis 2018;13:2577-2586 [FREE Full text] [doi: 10.2147/COPD.S170242] [Medline: 30214176]
13. Al Rajeh AM, Hurst JR. Monitoring of physiological parameters to predict exacerbations of chronic obstructive pulmonary disease (COPD): a systematic review. J Clin Med 2016 Nov 25;5(12):108 [FREE Full text] [doi: 10.3390/jcm5120108] [Medline: 27897995]
14. Tomasic I, Tomasic N, Trobec R, Krpan M, Kelava T. Continuous remote monitoring of COPD patients: justification and explanation of the requirements and a survey of the available technologies. Med Biol Eng Comput 2018 Apr;56(4):547-569 [FREE Full text] [doi: 10.1007/s11517-018-1798-z] [Medline: 29504070]
15. Buekers J, Theunis J, De Boever P, Vaes AW, Koopman M, Janssen EV, et al. Wearable finger pulse oximetry for continuous oxygen saturation measurements during daily home routines of patients with chronic obstructive pulmonary disease (COPD) over one week: observational study. JMIR Mhealth Uhealth 2019 Jun 06;7(6):e12866 [FREE Full text] [doi: 10.2196/12866] [Medline: 31199331]
16. Vegesna A, Tran M, Angelaccio M, Arcona S. Remote patient monitoring via non-invasive digital technologies: a systematic review. Telemed J E Health 2017 Jan;23(1):3-17 [FREE Full text] [doi: 10.1089/tmj.2016.0051] [Medline: 27116181]
17. Guarascio AJ, Ray SM, Finch CK, Self TH. The clinical and economic burden of chronic obstructive pulmonary disease in the USA. Clinicoecon Outcomes Res 2013;5:235-245 [FREE Full text] [doi: 10.2147/CEOR.S34321] [Medline: 23818799]
18. Lazar A. Why we use and abandon smart devices. 2015 Presented at: Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; Osaka. [doi: 10.1145/2750858.2804288]
19. Lee J, Kim D, Ryoo H, Shin B. Sustainable wearables: wearable technology for enhancing the quality of human life. Sustainability 2016 May 11;8(5):466. [doi: 10.3390/su8050466]
20. Hermsen S, Moons J, Kerkhof P, Wiekens C. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]
21. Medical Device Depot. MIR SmartOne Portable Spirometer URL: https://www.medicaldevicedepot.com/MIR-SmartOne-Portable-Spirometer-p/SmartOne.htm [accessed 2020-04-13]

22. CardiacDirect. MIR SpiroLab Light Spirometer URL: https://www.cardiacdirect.com/Category/spirometers-1.html [accessed 2020-04-13]
23. FL400 Fingertip Pulse Oximeter. URL: https://www.pulseoximeter.org/fl400-black.html [accessed 2020-04-13]
24. CardiacDirect. DPM 2 Pulse Oximeter with Charging Stand and Rechargeable Battery URL: https://www.cardiacdirect.com/Product/DPM-2-Pulse-Oximeter-with-Charging-Stand-and-Rechargeable-Battery.html [accessed 2020-04-13]
25. Propeller Health. URL: https://www.propellerhealth.com/ [accessed 2020-04-13]
26. Chen J, Kaye L, Tuffli M, Barrett MA, Jones-Ford S, Shenouda T, et al. Passive monitoring of short-acting beta-agonist use via digital platform in patients with chronic obstructive pulmonary disease: quality improvement retrospective analysis. JMIR Form Res 2019 Oct 23;3(4):e13286 [FREE Full text] [doi: 10.2196/13286] [Medline: 31647471]
27. Alshabani K, Attaway AA, Smith MJ, Majumdar U, Rice R, Han X, et al. Electronic inhaler monitoring and healthcare utilization in chronic obstructive pulmonary disease. J Telemed Telecare 2019 May 16:1357633X19850404. [doi: 10.1177/1357633X19850404] [Medline: 31096842]
28. Wagner A. PBS NewsHour. 2017 May 10. Asthma patients breathe easier with new bluetooth inhalers URL: https://www.pbs.org/newshour/health/asthma-patients-breathe-easier-new-bluetooth-inhalers [accessed 2020-04-13]
29. Cohero Health. URL: https://coherohealth.com/ [accessed 2020-04-13]
30. Melvin E, Cushing A, Tam A, Kitada R, Manice M. Assessing the use of BreatheSmart® mobile technology in adult patients with asthma: a remote observational study. BMJ Open Respir Res 2017;4(1):e000204 [FREE Full text] [doi: 10.1136/bmjresp-2017-000204] [Medline: 29071077]
31. Dickey, MR. TechCrunch. 2015 Sep 21. Cohero Health aims to improve patient care for the 26 million american asthma sufferers URL: https://tinyurl.com/y9z9w8pr [accessed 2020-04-13]
32. Spry Health. URL: https://spryhealth.com/ [accessed 2020-04-13]
33. Omron Health HeartGuide. 2019 Jun 21. URL: https://tinyurl.com/yab663sx [accessed 2020-04-13]
34. Garcia-Rio F, Rojo B, Casitas R, Lores V, Madero R, Romero D, et al. Prognostic value of the objective measurement of daily physical activity in patients with COPD. Chest 2012 Aug;142(2):338-346. [doi: 10.1378/chest.11-2014] [Medline: 22281798]
35. Shorofsky M, Bourbeau J, Kimoff J, Jen R, Malhotra A, Ayas N, Canadian Respiratory Research Network, CanCOLD Collaborative Research group. Impaired sleep quality in COPD is associated with exacerbations: the CanCOLD Cohort Study. Chest 2019 Nov;156(5):852-863. [doi: 10.1016/j.chest.2019.04.132] [Medline: 31150638]
36. Holt M, Yule B. Ambulatory monitoring of respiratory effort using a clothing-adhered biosensor. 2018 Presented at: Proceedings of the IEEE International Symposium on Medical Measurements and Applications; 2018; Rome p. 1-6. [doi: 10.1109/memea.2018.8438678]
37. SpireHealth. URL: https://spirehealth.com/ [accessed 2020-04-13]
38. Cosinuss One. URL: https://www.cosinuss.com/ [accessed 2020-04-13]
39. Current Health. URL: https://currenthealth.com/ [accessed 2020-04-13]
40. Health Care Originals. Adamm RSM URL: http://healthcareoriginals.com/solutions/#adamm-rsm [accessed 2020-04-13]
41. Roossien CC, Heus R, Reneman MF, Verkerke GJ. Monitoring core temperature of firefighters to validate a wearable non-invasive core thermometer in different types of protective clothing: concurrent in-vivo validation. Appl Ergon 2020 Feb;83:103001. [doi: 10.1016/j.apergo.2019.103001] [Medline: 31739139]
42. Passler S, Müller N, Senner V. In-ear pulse rate measurement: a valid alternative to heart rate derived from electrocardiography? Sensors (Basel) 2019 Aug 21;19(17):3641 [FREE Full text] [doi: 10.3390/s19173641] [Medline: 31438600]
43. Chaglla ES, Celik N, Balachandran W. Measurement of core body temperature using graphene-inked infrared thermopile sensor. Sensors (Basel) 2018 Oct 03;18(10):3315 [FREE Full text] [doi: 10.3390/s18103315] [Medline: 30282932]
44. Hegewald MJ, Gallo HM, Wilson EL. Accuracy and quality of spirometry in primary care offices. Ann Am Thorac Soc 2016 Dec;13(12):2119-2124. [doi: 10.1513/AnnalsATS.201605-418OC] [Medline: 27598295]
45. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132. [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
46. Lipnick MS, Feiner JR, Au P, Bernstein M, Bickler PE. The accuracy of 6 inexpensive pulse oximeters not cleared by the Food and Drug Administration: the possible global public health implications. Anesth Analg 2016 Aug;123(2):338-345. [doi: 10.1213/ANE.0000000000001300] [Medline: 27089002]
47. Walker PP, Pompilio PP, Zanaboni P, Bergmo TS, Prikk K, Malinovschi A, et al. Telemonitoring in chronic obstructive pulmonary disease (CHROMED): a randomized clinical trial. Am J Respir Crit Care Med 2018 Sep 01;198(5):620-628. [doi: 10.1164/rccm.201712-2404OC] [Medline: 29557669]
48. Ghosh S, Anderson WH, Putcha N, Han MK, Curtis JL, Criner GJ, Current and former investigators of the SPIROMICS sites and reading centers. Alignment of inhaled chronic obstructive pulmonary disease therapies with published strategies. analysis of the global initiative for chronic obstructive lung disease recommendations in SPIROMICS. Ann Am Thorac Soc 2019 Feb;16(2):200-208. [doi: 10.1513/AnnalsATS.201804-283OC] [Medline: 30216731]

## Abbreviations

**COPD:** chronic obstructive pulmonary disease
**HR:** heart rate
**ISO:** International Organization for Standardization
**RPM:** remote patient monitoring
**RR:** respiratory rate

*Edited by G Eysenbach; submitted 05.09.19; peer-reviewed by C Lin, SGS Shah, C Velardo; comments to author 09.12.19; revised version received 11.02.20; accepted 23.03.20; published 21.05.20.*

*Please cite as:*
*Fan KG, Mandel J, Agnihotri P, Tai-Seale M*
*Remote Patient Monitoring Technologies for Predicting Chronic Obstructive Pulmonary Disease Exacerbations: Review and Comparison*
*JMIR Mhealth Uhealth 2020;8(5):e16147*
*URL: http://mhealth.jmir.org/2020/5/e16147/*
*doi:10.2196/16147*
*PMID:32348262*

Review

# Mobile Phone–Based Behavioral Interventions in Pregnancy to Promote Maternal and Fetal Health in High-Income Countries: Systematic Review

Tasmeen Hussain[1], MD, MPH; Patricia Smith[2], MLIS; Lynn M Yee[2], MD, MPH

[1]Department of Internal Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

[2]Division of Maternal Fetal Medicine, Department of Obstetrics and Gynecology, Northwestern University Feinberg School of Medicine, Chicago, IL, United States

**Corresponding Author:**
Lynn M Yee, MD, MPH
Division of Maternal Fetal Medicine
Department of Obstetrics and Gynecology
Northwestern University Feinberg School of Medicine
250 E. Superior Street, #5-2145
Chicago, IL
United States
Phone: 1 312 472 4685
Email: lynn.yee@northwestern.edu

## Abstract

**Background:** Chronic diseases have recently had an increasing effect on maternal-fetal health, especially in high-income countries. However, there remains a lack of discussion regarding health management with technological approaches, including mobile health (mHealth) interventions.

**Objective:** This study aimed to systematically evaluate mHealth interventions used in pregnancy in high-income countries and their effects on maternal health behaviors and maternal-fetal health outcomes.

**Methods:** This systematic review identified studies published between January 1, 2000, and November 30, 2018, in MEDLINE via PubMed, Cochrane Library, EMBASE, CINAHL, PsycINFO, Web of Science, and gray literature. Studies were eligible for inclusion if they included only pregnant women in high-income countries and evaluated stand-alone mobile phone interventions intended to promote healthy maternal beliefs, behaviors, and/or maternal-fetal health outcomes. Two researchers independently reviewed and categorized aspects of full-text articles, including source, study design, intervention and control, duration, participant age, attrition rate, main outcomes, and risk of bias. Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines were followed, and the study was registered in PROSPERO before initiation.

**Results:** Of the 2225 records examined, 28 studies were included and categorized into 4 themes: (1) gestational weight gain, obesity and physical activity (n=9); (2) smoking cessation (n=9); (3) influenza vaccination (n=2); and (4) general prenatal health, preventive strategies, and miscellaneous topics (n=8). Reported sample sizes ranged from 16 to 5243 with a median of 91. Most studies were performed in the United States (18/28, 64%) and were randomized controlled trials (21/28, 75%). All participants in the included studies were pregnant at the time of study initiation. Overall, 14% (4/28) of studies showed association between intervention use and improved health outcomes; all 4 studies focused on healthy gestational weight. Among those, 3 studies showed intervention use was associated with less overall gestational weight gain. These 3 studies involved interventions with text messaging or an app in combination with another communication strategy (Facebook or email). Regarding smoking cessation, influenza vaccination, and miscellaneous topics, there was some evidence of positive effects on health behaviors and beliefs, but very limited correlation with improved health outcomes. Data and interventions were heterogeneous, precluding a meta-analysis.

**Conclusions:** In high-income countries, utilization of mobile phone–based health behavior interventions in pregnancy demonstrates some correlation with positive beliefs, behaviors, and health outcomes. More effective interventions are multimodal in terms of features and tend to focus on healthy gestational weight gain.

***(JMIR Mhealth Uhealth 2020;8(5):e15111)*** doi:10.2196/15111

**KEYWORDS**

mHealth; mobile health; pregnancy; smartphone; text messaging; mobile applications; software; chronic disease; health behavior

# Introduction

## Background

Pregnancy and the postpartum period are times of rapid medical, social, and behavioral changes for women and their families. This period is perceived to be a *window of opportunity* for health interventions because many women have enhanced access to health care during pregnancy and may have increased motivation to improve their health during this time. Healthy maternal behaviors have been shown to improve the risk of pregnancy-related morbidities [1]. For example, smoking cessation, exercise, and healthy weight gain in pregnancy have all been linked to better maternal and fetal health [2-4].

Chronic disease is a particularly important arena. Per the Centers for Disease Control and Prevention, although the rate of maternal death related to traditional risk factors such as hemorrhage, hypertensive disorders of pregnancy, and anesthesia complications in the United States is decreasing, mortality related to cardiovascular disease (CVD), cerebrovascular accidents, and other medical conditions continues to increase [5]. Cardiovascular conditions were responsible for more than one-third of all pregnancy-related deaths in the United States between 2011 and 2016. Thus, pregnancy is an important time to improve health behaviors, such as promoting healthy gestational weight gain and managing chronic disease. However, changes to health behaviors often require intensive provider support, consistent follow-up, and frequent counseling that are difficult to maintain during short outpatient visits. These requirements may be supported by technology.

Many health behavior and lifestyle interventions have incorporated technology in various areas of chronic disease management [6,7]. In particular, the field of mobile health (mHealth) has recently seen rapid growth. mHealth refers to the use of mobile technologies including mobile phones, personal digital assistants, and even tablet computers to improve patient health. Outside of pregnancy, a growing amount of literature suggests that mHealth and other digital interventions are feasible, acceptable, and may promote improved health behaviors [8-10]. An estimated 76% of people in high-income countries own a mobile phone, and 87% use the internet [11,12]. Furthermore, a more focused study of pregnant women in the United States showed that 88% had access to a mobile phone, and 89% had access to the internet [13]. These data suggest both are promising media for use with pregnant women in the management of chronic conditions in high-income environments.

Past studies suggest women are interested in receiving health information on the Web and are comfortable with using their mobile phones [14]. However, research on the use of mHealth in pregnancy has been broad and heterogeneous. Much of the research done is with small groups in low- or middle-income countries or utilizes a *telemedicine* format, defined as technology-facilitated direct communication with medical professionals [15-22]. There remains a lack of organized discussion on mHealth interventions in pregnancy that are tailored or self-maintaining as well as on studies of women in high-income countries, where access to mobile phones is the greatest and women are highly affected by chronic disease

## Objective

The objective of this study was to systematically evaluate mHealth interventions used in pregnancy in high-income countries and their effects on maternal health behaviors and maternal-fetal health outcomes.

# Methods

## Study Registration

Before performance of this search, information about the study proposal was published electronically in the University of York PROSPERO register of systematic reviews [23]. The authors followed all guidelines for the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [24].

## Eligibility Criteria, Information Sources, and Search Strategy

We conducted a systematic review of studies on mobile phone–based mHealth interventions designed for pregnant women. A research librarian (PS) was primarily responsible for a comprehensive literature search. We included English-language articles with a patient population that included pregnant women who utilized pregnancy-related mobile phone interventions during their pregnancy. In addition, we limited our studies to those performed in developed or high-income countries as defined by The World Economic Situation and Prospects 2012 of the United Nations [25]. Study types included meta-analyses, systematic reviews, randomized controlled trials (RCTs) including randomized crossover trials and cluster randomized trials, and nonexperimental observational studies. The assessed interventions were stand-alone mobile phone–based interventions including, but not limited to, mobile phone apps, text messaging, games, and information services. We excluded studies that used technology interventions aimed solely at communication between patients and clinicians without a stand-alone educational, motivational, or interactive component (such as telemedicine portals or electronic medical record–based portals for use with mobile phones), and interventions that were not primarily intended for mobile phone use, eg, websites. Studies were excluded if they focused solely on neonatal health, such as neonatal feeding support interventions or growth tracking tools. Studies were also excluded if they were exclusively published as abstracts or conference proceedings without a full peer-reviewed manuscript. Finally, studies were excluded if they were solely meant to evaluate feasibility or desirability of hypothetical interventions or supplied outcomes with fewer than 2 weeks of intervention use.

We searched MEDLINE via PubMed, EMBASE, Web of Science, Cochrane Database of Controlled Trials, CINAHL,

and PsycINFO databases from January 1, 2000 to November 30, 2018. We began with the MEDLINE search and translated to the appropriate syntax for each of the other databases, using controlled vocabulary when possible. Search terms related to pregnancy, mobile interventions, and select behaviors (including smoking cessation, weight loss, and diabetes management) were included. Full search strategies can be found in Multimedia Appendix 1, and a completed PRISMA checklist is found in Multimedia Appendix 2.

### Study Selection

Titles and abstracts of studies were read by 2 independent reviewers (TH and LY) on two online abstract organizers (abstrackr: [26] and Rayyan [27]). Discordant assessments were resolved by discussion between reviewers or with the involvement of a third author (PS) when necessary.

Studies were then divided into 4 subgroups based on their primary clinical focus: (1) gestational weight gain, obesity and physical activity; (2) smoking cessation; (3) influenza vaccination; and (4) general prenatal health, preventive strategies, and miscellaneous topics. For all study types, data extraction was standardized to include source, study design, number of participants in the intervention and control groups, intervention and control descriptions, duration, participant age and other details if available, attrition rates, and main outcomes.

### Data Extraction

Two authors (TH and LY) simultaneously reviewed all abstracts for inclusion using Abstrackr and Rayyan, as described above. EndNote X7.2 (EndNote, Clarivate Analytics, Philadelphia, Pennsylvania, United States) was used to identify and remove duplicate records. Two searches were conducted; the initial search reviewed literature to 2016 and an updated search reviewed more recent literature until November 30, 2018. Once relevant abstracts were agreed upon, full-text analysis of included abstracts was then performed by the same authors. In addition, review of the bibliographies of included full-text articles were reviewed for additional eligible articles. Relevant articles meeting the final inclusion criteria were then abstracted in-depth for bias, study quality, and overall findings.

### Assessment of Risk of Bias

Bias was evaluated by 2 independent reviewers (TH and LY). We applied specific tools for assessment of risk of bias tailored to each study type. For observational studies (not randomized controls), we used 1 of 2 National Institutes of Health (NIH) Quality Assessment Tools (NIH QAT), which consisted of 12 items to assist raters in formulating a holistic final quality assessment [28]. If a study had a control but was not randomized, the Quality Assessment of Controlled Intervention Studies was used; if no control was available, the Quality Assessment Tool for Before-After (Pre-Post) Studies with No Control Group was used. For RCTs, we used the Cochrane Risk of Bias tool [29]. Studies were rated independently by 2 reviewers (TH and LY). Disagreements were resolved by discussion between reviewers or with the involvement of a third author (PS) when necessary.

### Data Synthesis

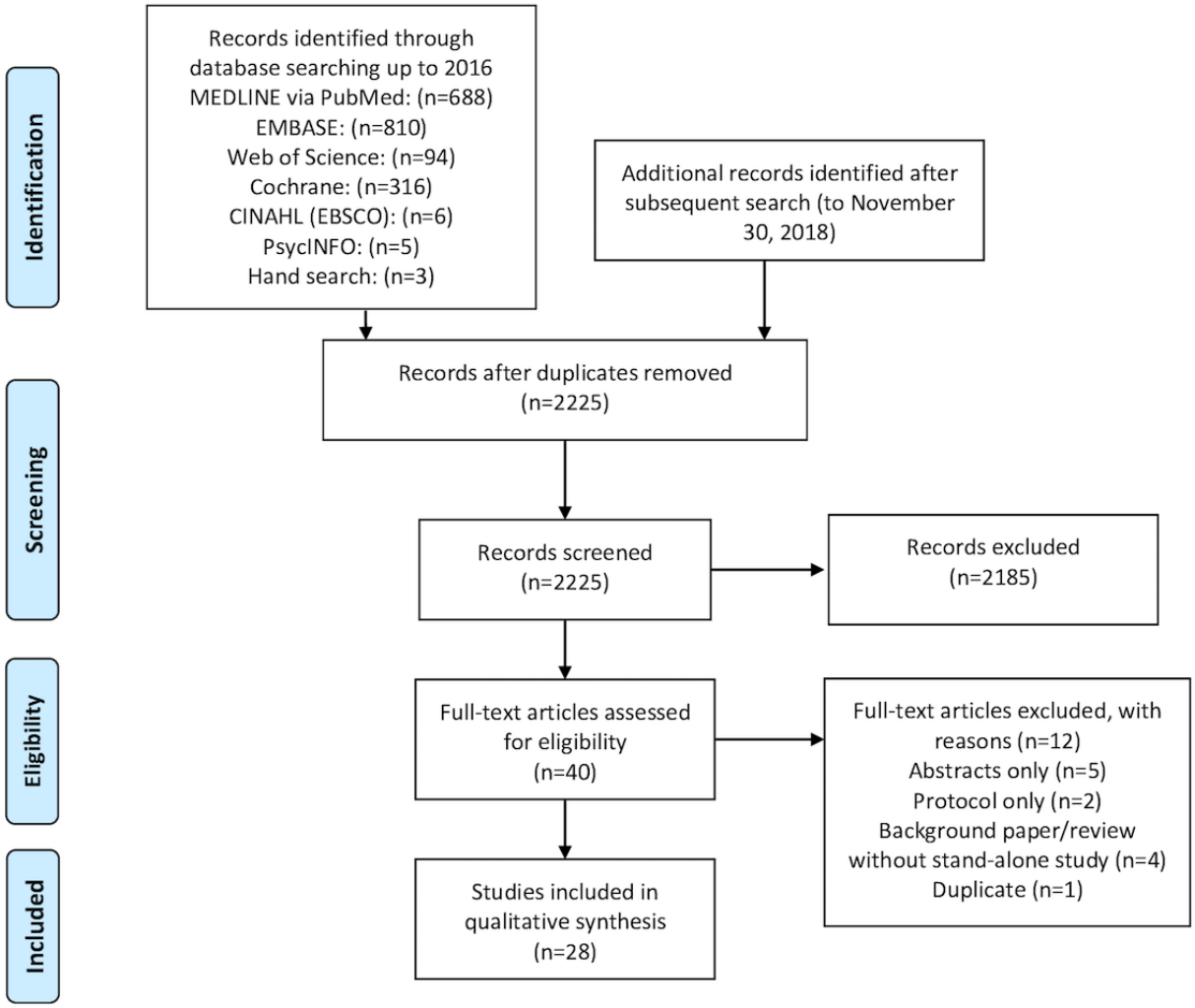
Data were collected to be primarily presented descriptively. We considered a meta-analysis or pooling of data if sufficient homogeneity in measured outcomes were to be observed, but the evaluation of data demonstrated heterogeneity that precluded such analyses.

## Results

### Study Selection

An electronic search as described previously revealed a total of 2225 titles and abstracts after the removal of duplicates. After full-text evaluation, a total of 28 studies met the criteria for inclusion. An adapted PRISMA study flowchart is shown in Figure 1.

**Figure 1.** Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow sheet.




## Study Characteristics and Synthesis of Results

Included studies fell into 4 categories: (1) gestational weight gain, obesity, and physical activity (n=9); (2) smoking cessation (n=9); (3) influenza vaccination (n=2); and (4) general prenatal health, preventive strategies, and miscellaneous topics (n=8). Reported sample sizes ranged from 16 to 5243 with a median of 91. All participants in the included studies were pregnant at the time of study initiation.

Tables 1 and 2 outline studies focused on gestational weight gain, obesity, and physical activity [28,30-37]. Of the 9 eligible studies, 2 used exclusively text messages, 4 utilized text messages in conjunction with other technology, and 3 utilized mobile phone apps without text messages. One included study was not randomized, whereas the remainder were RCTs. Outcomes varied widely among studies. Two studies showed that intervention participants were significantly less likely to exceed healthy gestational weight gain during pregnancy (37% vs 66%; *P*=.03) [31], and (58% vs 85%; *P*=.04) [36], but one found no such difference [33]. Three interventions were associated with less overall gestational weight gain in intervention users over the study period [31,33,37]. Notably, each of these interventions were multimodal and incorporated at least one additional communicative technology (Facebook or emails) alongside its main intervention (text messages or an app). The 2 studies that evaluated gestational weight gain and utilized interactive text messages or an app alone exhibited no difference in gestational weight gain compared with controls [34,36]. Studies also differed regarding behavior change. Although some showed improvements in behavior analogs such as increased self-reported exercise [37] and less reduction in physical activity during pregnancy compared with prepregnancy [33], others showed no such relationship [32,34,35]. There was no difference in incidence of gestational diabetes in any study [31,37]. In terms of cost, 1 study did find that a mobile app compared with a parallel intervention requiring in-person counseling by health coaches was significantly less expensive (US $97 vs US $347) [36]. In this study, both remote and in-person interventions were associated with lower proportion of excess gestational weight gain when compared with controls.

**Table 1.** Design of trials with a focus on gestational weight gain, obesity, and physical activity.

| Reference | Setting/country/population | Study design | Experimental arm vs control arm(s), n | Intervention description and control | Duration |
|---|---|---|---|---|---|
| Soltani et al (2015) [28] | Prenatal clinic/Doncaster, England/BMI>30; 8-10 weeks' gestation | Observational | Intervention vs control: 16 vs 15 | MOMTech text messages: 14 motivational text messages per week, food and activity diary, goal setting, and consultation visits vs usual care | Until 6 weeks postpartum |
| Choi et al (2016) [30] | Prenatal clinics and community/San Francisco, CA, United States/Sedentary; 10-20 weeks' gestation | RCT[a] | Intervention vs control: 15 vs 15 | Mobile app: Fitbit-enhanced daily message as text message or short video script, activity diary, and automated feedback vs Fitbit only | 12 weeks |
| Herring et al (2016) [31] | Prenatal clinics/Philadelphia, Pennsylvania, United States/African American; <20 weeks' gestation; BMI 25-45 | RCT | Intervention vs control: 33 vs 33 | Behavior change goals, interactive self-monitoring text messages, biweekly health coach calls, and skills training and support through Facebook vs usual care | Until 36 weeks' gestation |
| Dodd (2017) [32] | Public maternity hospitals/Adelaide, South Australia/10-20 weeks' gestation | RCT | Intervention vs control: 77 vs 85 | Interactive mobile phone app with information about dietary guidelines and physical activity guidelines during pregnancy; also encouraged women to set dietary and physical activity goals and monitor their progress vs lifestyle advice only | Until 36 weeks' gestation |
| Willcox (2017) [33] | Academic maternity hospital/Melbourne, Australia/10-17 weeks' gestation; prepregnancy BMI>25 | RCT | Intervention vs control: 45 vs 46 | txt4two: Tailored text messages, Web-based app, video messages, and Facebook chat room and brochure vs brochure only | Until 36 weeks' gestation |
| Pollak (2014) [34] | Prenatal clinics/Durham, NC, United States/prepregnancy BMI=25-40; 12-21 weeks' gestation | RCT | Intervention vs text4baby: 22 vs 11 | Preg CHAT texts: interactive 3 times weekly texts regarding behaviors–step counts, sweetened drinks, fruits/vegetables, and eliminating fast foods vs text4baby alone | 16 weeks |
| Huberty (2017) [35] | Online/US residents/8-16 weeks' gestation; low physical activity | RCT | 3 intervention groups: Plus One group (21); Plus Six (20); Plus Six Choice (18); and standard group (21) | 3 intervention groups with variations on general and physical activity texts received per week: Plus One, Plus Six, Plus Six Choice; participants also received Fitbit flex to track sleep and exercise data. All were compared with the *standard* group, which was three text4baby SMS per week at noon | Until 40 weeks' gestation |
| Redman (2017) [36] | Various clinics/United States/BMI=25-39.9; first trimester of pregnancy | RCT | 2 intervention groups: In person (18), Remote (19), and control (17) | 2 intervention groups: In person—dietary intake advice, exercise advice, paper weight graph and counseling provided by health coaches; Remote—same information as above provided in a mobile app format with electronic data capture; both compared with usual care from obstetrician | Until delivery |
| Kennelly (2018) [37] | Maternity hospital/Dublin, Ireland/BMI=25-39.9; 10-15 weeks' gestation | RCT | Intervention vs usual care: 278 vs 287 | A mobile phone app with low glycemic index recipes, an exercise advice section, and a home page with tips and encouraging thought of the day. Also received emails every 2 weeks and two face-to-face hospital visits vs usual care | Until 34 weeks' gestation |

[a]RCT: randomized controlled trial.

**Table 2.** Outcomes and bias of trials with a focus on gestational weight gain, obesity, and physical activity.

| Reference | Participant age (years), mean (SD) | Attrition rate | Main outcomes | Bias tool | Bias rating | Bias reasoning |
|---|---|---|---|---|---|---|
| Soltani et al (2015) [28] | 29.1 (5.4) for IG[a] vs 31.7 (5.8) for CG[b] | 13% (2/16) | • No significant difference in mean GWG[c] (5.6 vs 9.7 kg)<br>• No significant difference in percentage of participants who exceeded the IOM[d] upper limit of GWG for obese women (28% vs 50%) | NIH QAT[e] | Fair risk | Small sample size |
| Choi et al (2016) [30] | 32.9 (2.5) for IG vs 34.5 (2.5) in CG | 40% to daily messages, 33% to activity diary | • Significantly less "Lack of energy as a barrier to being active," at week 12 in IG ($P$=.02)<br>• No difference between groups in change in weekly mean steps ($P$=.23)<br>• No change in numerous outcomes including CES-D[f] score, severity of pregnancy symptoms, self-efficacy | Cochrane ROBT[g] | Low risk | N/A[h] |
| Herring et al (2016) [31] | 25.9 (4.9) for IG vs 25.0 (5.7) for CG | Unclear | • Significantly greater percentage of IG kept within IOM guidelines for GWG (37% vs 66%; $P$=.03)<br>• Significant adjusted mean difference in total GWG in IG, early pregnancy to delivery (8.7 vs 12.3 kg; $P$=.046)<br>• No significant difference in mean birth weight or babies small or large for gestational age. No difference in percentage of women with GDM[i] | Cochrane ROBT | Low risk | N/A |
| Dodd (2017) [32] | 30.87 (5.07) for IG vs 31.01 (6.16) for CG | 38.2% (62/162) | • No significant difference in self-reported Healthy Eating Index scores, macronutrient and food group intake, or physical activity | Cochrane ROBT | High risk | High attrition, self-report, and women knew allocations |
| Willcox (2017) [33] | 33.0 (3.4) for IG vs 32.0 (5.1) for CG | 9.0% (9/100) | • Significantly less GWG with txt4two (7.8 vs 9.7 kg; adjusted $P$=.04)<br>• Significantly fewer txt4two women reduced their minutes of total daily physical activity over the course of the intervention ($P$=.001)<br>• No significant difference in proportion of women exceeding IOM GWG guidelines. (47% vs 61%; adjusted $P$=.07)<br>• No significant differences in self-reported consumption of food groups | Cochrane ROBT | High risk | Women not blinded, self-reported exercise |
| Pollak (2014) [34] | 29 (5) for IG vs 32 (2) in CG | 30% (10/33) | • No significant difference in mean weight gain, physical activity level outcomes, or nutrition score | Cochrane ROBT | High risk | High proportional attrition, low sample size. Possibly randomized by study staff |
| Huberty (2017) [35] | 31.05 (5.52) for Plus One vs 31.48 (5.44) for Plus Six vs 31.44 (4.16) for Plus Six Choice vs 30.83 (5.22) for standard | 14% (13/93) | • All 3 IGs were consolidated; when compared with controls, no difference in linear trajectories or quadratic trajectories regarding active time, light intensity time, and steps | Cochrane ROBT | Fair risk | Not blinded |

| Reference | Participant age (years), mean (SD) | Attrition rate | Main outcomes | Bias tool | Bias rating | Bias reasoning |
|---|---|---|---|---|---|---|
| Redman (2017) [36] | 29.0 (4.2) for remote vs 29.2 (4.8) for in person vs 29.5 (5.1) for CG | Unclear | • Significantly lower proportion of women with excess GWG in the remote group compared with usual care groups (58% vs 85%; *P*=.04)<br>• No significant difference in GWG between the remote group and usual care (least squares mean 10.0 vs 12.8 kg; *P*=.07)<br>• Significantly less intervention cost for remote compared with in-person group (US $97 vs US $347; *P*<.001) | Cochrane ROBT | High risk | Randomized by unblinded intervention staff |
| Kennelly (2018) [37] | 32.8 (4.6) for IG vs 32.1 (4.2) for CG | 11.9% (67/565) | • No significant difference in incidence of GDM (15.4% vs 14.1%; *P*=.71)<br>• Significantly less GWG in IG (8.9 vs 10 kg; *P*=.02)<br>• Significantly lower dietary glycemic load (*P*=.02) and increased exercise in IG (*P*=.02) after multiple correction testing | Cochrane ROBT | Fair risk | Self-reported exercise and food outcomes; neither participants nor researchers blinded |

[a]IG: intervention group.

[b]CG: control group.

[c]GWG: gestational weight gain.

[d]IOM: Institute of Medicine.

[e]NIH QAT: National Institutes of Health Quality Assessment Tool.

[f]CES-D: Center for Epidemiologic Studies Depression Scale.

[g]N/A: not applicable.

[h]ROBT: risk of bias tool.

[i]GDM: gestational diabetes mellitus.

Tables 3 and 4 outline interventions to address smoking cessation during pregnancy [38-46]. Of the 9 studies, 7 used exclusively text messages, and the remaining 2 studies used mobile phone apps. Overall, outcomes were sparse regarding the ability of interventions to affect smoking cessation. Although 2 small uncontrolled studies showed a decrease in cigarettes smoked over the course of intervention [38] and more than 70% achievement of nonsmoking by the end of the intervention [39], the studies that employed control arms showed no difference in outcomes. These outcomes varied but included self-reported abstinence, biochemically reported abstinence, and number of smoke-free days [40-46]. In 1 study, using text messages as the intervention mode was associated with increased self-efficacy, determination to quit smoking in pregnancy, and setting a quit date [42].

Tables 5 and 6 highlight the 2 studies of interventions to improve influenza vaccination rates [47,48]. Both utilized text messages alone. There was no difference in influenza vaccination rates in intervention vs control groups in either study.

Tables 7 and 8 outline the remaining 8 studies, which focused on general prenatal health, preventive strategies, and miscellaneous topics [49-56]. Four studies employed text messages alone, and 4 used mobile phone apps. In this sphere, interventions were most associated with improvements in health beliefs [49] and behaviors including self-reported attempts to eat more nutritious food [50], belief that taking prenatal vitamins will improve the health of the fetus [54], and belief that the participant is prepared to be a new mother [54]. There was also a significant association between intervention use and attending a prenatal visit at least 6 months before delivery in 1 controlled study [51]. In 1 study without formal controls, there was a higher rate of clinic attendance in intervention users (84%) compared with that for the general clinic population (50%) [52]. In this study, attendance was even higher (89%) than in those who scheduled transportation through a free rideshare service facilitated through the app. Among these studies, there was no difference in any measured health outcomes including cesarean delivery and neonatal intensive care unit admission [51], hypertensive disorders of pregnancy, gestational weight gain, delivery outcomes [53], and beliefs and behaviors around smoking and alcohol [54,55]. One unique study employed a mobile phone app to improve rates of perineal massage in Japan; this intervention was not associated with any difference in rates of practice of perineal massage, perineal lacerations, or episiotomy rates [56].

**Table 3.** Design of trials with a focus on smoking cessation.

| Reference | Setting/country/population | Study design | Experimental arm vs control arm(s), n | Intervention description and control | Duration |
|---|---|---|---|---|---|
| Abroms et al (2015) [38] | Online/United States/current smoker or recently quit (<4 weeks ago), <30 weeks' gestation | Observational | Intervention (20), no control arm | Quit4baby text messages: 1-5 messages per day in reference to chosen quit date; also included interactive keyword-based support messages. Participants continued to receive text4baby messages concurrently | 4 weeks |
| Fujioka et al (2012) [39] | Obstetrics consultations/Yamaguchi prefecture, Japan/current smokers, >20 weeks' gestation | Observational | Intervention (52), no control arm | Mobile phone e-learning program: smoking cessation education, ability to set quit date, ability to select who will help quit smoking, record of declaration of quitting smoking | 3 months |
| Abroms (2017) [41] | Prenatal clinics/Washington DC, United States/current smoker or recently quit (<2 weeks ago) | RCT[a] | Intervention (55) vs control (44) | SmokefreeMOM: Tailored and interactive texts 3-6 times per day regarding smoking including setting a quit date, self-efficacy, and expectations regarding quitting vs usual care | 3 months |
| Naughton et al (2012) [42] | Prenatal clinics/England/current smokers, <21 weeks' gestation | RCT | Intervention (102) vs control (105) | MiQuit text messages: Tailored text messages 0-2 times/day at random intervals as well as *instant-response* supportive texts for help or lapses in behavior and tailored leaflet vs untailored leaflet | 3 months |
| Pollak et al (2013) [43] | Prenatal clinics/United States/current smokers, 10-30 weeks' gestation | RCT | Intervention (16) vs control (15) | Scheduled gradual reduction SMS: Gradual program to reduce smoking to 0 cigarettes by the 4th week. Support messages included up to 5 messages per day about various smoking cessation topics as well as setting a quit date vs support messages alone | 5 weeks |
| Tombor (2018) [44] | Online/England/current smokers | RCT | 565 randomized to one of 32 groups in full factorial design, randomized to *full* or *minimal* version of each module | SmokeFree Baby App assessed 5 modules: identity, health information, stress management, face-to-face support, behavioral substitution | 4 weeks |
| Abroms (2017) [40] | Online/United States; current smokers | RCT | Intervention (250) vs control (247) | Quit4baby: Tailored and interactive texts 1-8 times/day regarding smoking including setting a quit date, self-efficacy, and expectations regarding quitting. Was employed in addition to Text4baby. Compared with Text4baby alone | 3 months |
| Forinash (2018) [45] | Prenatal clinic, St. Louis, MO, United States/current smokers | RCT | Intervention (14) vs control (16) | Text messages every several days in a tapering pattern with encouragement to stop smoking vs usual care | 8 weeks |

| Reference | Setting/country/population | Study design | Experimental arm vs control arm(s), n | Intervention description and control | Duration |
|---|---|---|---|---|---|
| Naughton (2017) [46] | Prenatal clinics, England/<25 weeks' gestation; current smokers | RCT | Intervention (203) vs control (204) | MiQuit: an automated 12-week advice and support program for quitting smoking delivered by SMS text message. Tailored to desired themes including gestation, motivation to quit, self-efficacy, and partner's smoking status vs usual care | Until 36 weeks' gestation |

[a]RCT: randomized controlled trial.

**Table 4.** Outcomes and bias of trials with a focus on smoking cessation.

| Reference | Participant age (years), mean (SD) | Attrition rate | Main outcomes | Bias tool | Bias rating | Bias reasoning |
|---|---|---|---|---|---|---|
| Abroms et al (2015) [38] | 28.1 (6.1) for total sample | 35% (7/20) | • Cigarettes smoked decreased from 7.6 (4.9) to 2.4 (1.8) after 4 weeks but was not significant | NIH QAT[a] | Fair risk | No pre- to postanalysis, multiple measurements not taken, high loss to follow-up, high attrition |
| Fujioka et al (2012) [39] | 25.9 (4.7) for total sample | 7.7% (4/52) | • 71.1% of participants achieved nonsmoking<br>• Confidence to continue not smoking increased in both groups (those who ended up smoking, and those who quit smoking) | NIH QAT | Fair risk | Not all eligible participants were enrolled, measurements not taken multiple times |
| Abroms (2017) [41] | 27.18 (4.98) for IG[b] vs 28.25 (4.78) for CG[c] | 26% (26/99) | • No significant differences in any smoking-related outcomes including biochemically confirmed 7-day PPA[d], self-reported 7-day and 30-day abstinence, consecutive days quit, quit attempts, and changes in cigarettes smoked/day | Cochrane ROBT[e] | High risk | No information about blinding; randomization scheme changed in the middle of study |
| Naughton et al (2012) [42] | 27.2 (6.4) for IG vs 26.5 (6.2) for CG | 11% (23/207) | • Significantly higher overall self-efficacy, habitual self-efficacy, social self-efficacy and determination to quit smoking in pregnancy in IG<br>• Significantly higher probability to set a quit date in intervention group (45% vs 30%)<br>• No difference in outcomes including self-reported point prevalence at 3, 7, and 12 weeks, or making at least one 24 hour quit attempt | Cochrane ROBT | Low risk | N/A[f] |
| Pollak et al (2013) [43] | 29 (6) for IG and 27 (6) for CG | 6% (2/31) | • No change in 7-day point prevalence (7.5% vs 13.4%) or cigarettes smoked | Cochrane ROBT | High risk | Blinding and randomization strategies unclear |
| Tombor (2018) [44] | 27.3 (5.5) for total sample | 68.9% (389/565) | • No module was associated with fewer smoke-free days | Cochrane ROBT | High risk | Very high attrition rate; of note, all lost to follow-up were assumed to be smokers |
| Abroms (2017) [40] | 26.68 (5.94) for IG vs 25.95 (5.74) for CG | 28.2% (140/497) | • Overall, no significant difference in biochemically confirmed 7-day PPA at 3-month follow-up (39% vs 27%) IG vs CG | Cochrane ROBT | Fair risk | Self-reporting patients knew which group they were in |
| Forinash (2018) [45] | Not provided | 39% (19/49) | • No significant difference was found in eCO[g]-verified cessation (57.1% vs 31.3%; *P*=.15), eCO below 8 ppm at ≥1 visit (64.3% vs 37.5%; *P*=.14), or in birth outcomes | Cochrane ROBT | High risk | High attrition |
| Naughton (2017) [46] | 26.6 (5.7) for IG vs 6.4 (5.7) for CG | 35.9% (146/407) | • No difference in self-reported, later biochemically confirmed, abstinence in late pregnancy | Cochrane ROBT | Fair risk | Moderately high attrition; of note, those lost to follow-up assumed to be smokers |

[a]NIH QAT: NIH Quality Assessment Tool.

[b]IG: intervention group.

[c]CG: control group.

[d]PPA: point prevalence abstinence.

[e]ROBT: risk of bias tool.

[f]N/A: Not applicable.

[g]eCO: exhaled carbon monoxide.

**Table 5.** Design of trials with a focus on influenza vaccination.

| Reference | Setting/country/population | Study design | Experimental arm vs control arm(s), n | Intervention description and control | Duration |
|---|---|---|---|---|---|
| Moniz et al (2013) [47] | Prenatal clinic/Pittsburgh, Pennsylvania, United States/<28 weeks' gestation | RCT[a] | Intervention (104) vs control (100) | 12 weekly text messages with information about the benefits and safety of influenza vaccine in pregnancy vs control messages about general pregnancy health alone | 12 weeks, up to 6 weeks postpartum |
| Yudin (2017) [48] | Prenatal clinic/Toronto, Canada/any gestational age | RCT | Intervention (153) vs control (164) | Text messages twice weekly × 4 weeks emphasizing pregnant women's susceptibility to influenza, effectiveness of vaccine, safety of vaccines, and that vaccines are recommended for pregnant women vs usual care | Until 6 weeks postpartum |

[a]RCT: randomized controlled trial.

**Table 6.** Outcomes and bias of trials with a focus on influenza vaccination.

| Reference | Participant age (years), mean (SD) | Attrition rate | Main outcomes | Bias tool | Bias rating | Bias reasoning |
|---|---|---|---|---|---|---|
| Moniz et al (2013) [47] | Ranged 13-49 | 23% (46/204) | No difference in influenza vaccination rate (33% vs 31%) | Cochrane ROBT[a] | Low risk | N/A[b] |
| Yudin (2017) [48] | 32.2 (4.5) for IG[c] vs 32.4 (4.9) for CG[d] | 10.7% (34/317) | No difference in influenza vaccination rate (31% vs 27%; *P*=.51) | Cochrane ROBT | Low risk | N/A |

[a]ROBT: risk of bias tool.

[b]N/A: not applicable.

[c]IG: intervention group.

[d]CG: control group.

**Table 7.** Design of trials with a focus on general health, preventive health, health beliefs, and other topics.

| Reference | Topic | Setting/country/population | Study design | Experimental arm vs control arm(s), n | Intervention description and control | Duration |
|---|---|---|---|---|---|---|
| Moniz et al (2015) [49] | Preventive health behaviors (smoking cessation, condom use, nutrition optimization, seat belt use, breastfeeding) | Prenatal clinic/Pittsburgh, Pennsylvania, United States/<28 weeks' gestation | Observational | Intervention (171), no control arm | General preventive health text messages regarding tobacco cessation, sexually transmitted disease prevention, daily vitamin use, seat belt use, dietary discretion and breastfeeding | 12 weeks |
| Dalrymple et al (2013) [50] | General prenatal health topics | Prenatal clinic/Philadelphia, Pennsylvania, United States/no special population | Observational | Intervention (31), no control arm | Twice weekly text messages delivered alongside text4baby messages on days text4baby messages were not sent | Unclear |
| Bush (2017) [51] | Numerous including weight, milestones, Wyoming-specific resources | Online/Wyoming state, United States/Medicaid users | Observation | Intervention (85) vs non–app-Medicaid members (5158) | Wyhealth Due Date Plus: mobile phone app that includes information on 70 health risk factors, provides pregnancy timeline, weight tracker, and appointment reminders | 6 months |
| Krishnamurti (2017) [52] | Numerous, including nutrition, routine prenatal care, violence, smoking, preterm labor | Prenatal clinic/Pittsburgh, Pennsylvania, United States/Medicaid-qualifying women | Observation | Intervention (16), no control arm | My Healthy Pregnancy App: Interactive application that gathered data regarding risk factors and delivered patient-specific risk feedback and recommendations. Could also arrange for Uber rides to clinic | 3 months or until delivery |
| Ledford (2016) [53] | General obstetric care, health literacy | Prenatal clinic/Bethesda, MD, United States/10-12 weeks' gestation | RCT[a] | Intervention (87) vs control (86) | Mobile app for journaling with space for recording weight, blood pressure, and experience between prenatal appointments vs spiral notebook alone | Until 32 weeks' gestation |
| Evans et al (2014) [54] | General prenatal health topics | Army Medical Center/Tacoma, WA, United States/<14 weeks' gestation | RCT | Intervention (498) vs control (498) | Text messages: 3 automated, tailored text messages per week vs usual care | 4 weeks |
| Evans et al (2012) [55] | General prenatal health topics | Prenatal clinic/Fairfax county, VA, United States/largely low-income | RCT | Intervention (48) vs control (38) | Automated, tailored text messages vs usual care | 28 weeks' gestation |
| Takeuchi (2016) [56] | Perineal massage | Prenatal clinic/Tokyo, Japan/30-33 weeks' gestation | RCT | Intervention (81) vs control (80) | Mobile phone website underlining effects of perineal massage, massage technique, support through peer group, communication with professional, and reminders/encouragement vs leaflet alone | Until delivery |

[a]RCT: randomized controlled trial.

**Table 8.** Outcomes and bias of trials with a focus on general health, preventive health, health beliefs, and other topics.

| Reference | Participant age (years), mean (SD) | Attrition rate | Main outcomes | Bias tool | Bias rating | Bias reasoning |
|---|---|---|---|---|---|---|
| Moniz et al (2015) [49] | 24.0 (4.5) | 8% (13/171) | Participants agreed that receiving text messages changed their beliefs about targeted preventive health behaviors:<br>• Smoking (50%)<br>• Sexually transmitted disease prevention (72%)<br>• Prenatal vitamins (83%)<br>• Seat belt use (68%)<br>• Nutritious food intake (84%)<br>• Breastfeeding (68%) | NIH QAT[a] | Fair risk | No before/after or multiple measurements taken |
| Dalrymple et al (2013) [50] | Unclear | 84% (26/31) for posttest; 35% (11/31) for any monthly form | • 100% agreed "I tried to eat better for myself and the baby."<br>• 60% agreed "I understood what was happening to my body better." | NIH QAT | High risk | No before/after or multiple measurements taken, small sample size, high attrition |
| Bush (2017) [51] | Unclear | Unclear | • Significant association between app use and completion of a prenatal visit at least 6 months before delivery (OR[b] 1.76; *P*=.02)<br>• Borderline significant association between app use and low birth weight (OR 0.25; *P*=.06)<br>• No association between app use and cesarean delivery or NICU[c] admission | NIH QAT | High risk | Used a comparison that was not randomly selected (self-selected app users) |
| Krishnamurti (2017) [52] | Median 24, range (18-35) | 0% (0/16) | • Intervention users reported higher intention to breastfeed at 2 months ($t_{13}$=−4.16; *P*=.001) and 3 months ($t_{15}$=−2.76; *P*=.01)<br>• No statistical significance in intention to use prenatal vitamins<br>• Clinic attendance rate was higher in participants than nonparticipant clinic patients (84% vs 50%)<br>• Attendance was even higher (89%) among those who scheduled free Uber transportation | NIH QAT | High risk | Sample size too low |
| Ledford (2016) [53] | 29.29 (4.8) for IG[d] vs 29.37 (4.83) for CG[e] | 27% (46/173) | • Mobile group reported more frequent use (*P*=.04) and greater patient activation (*P*=.02) than the notebook group<br>• No difference in biometrics including blood pressure control, weight gain, delivery outcomes | Cochrane ROBT[f] | Fair risk | Unclear how randomization occurred, patients not blinded |
| Evans et al (2014) [54] | 26.53 (SD not noted) | 51.3% (484/943) | • Significantly more of the intervention group agreed that "Taking prenatal vitamins will improve the health of my developing baby" (OR 1.91; *P*=.02)<br>• No difference in outcomes including self-reported smoking, consumption of alcoholic beverages or fruit and vegetable consumption | Cochrane ROBT | Fair risk | Selective reporting, high attrition |
| Evans et al (2012) [55] | 27.6 (SD not noted) | 27% (33/123) | • Significantly more of the intervention group agreed that "I am prepared to be a new mother" (OR 2.73; *P*=.04)<br>• No difference in outcomes including beliefs that smoking will harm the developing baby, that drinking alcohol will harm the developing baby, and that taking prenatal vitamins will improve the health of the developing baby | Cochrane ROBT | Fair risk | Unclear blinding of participants and personnel; incomplete outcome data |
| Takeuchi (2016) [56] | 32.7 (4.59) for IG vs 32.5 (4.18) for CG | 40% (65/161) | • No difference in practice of perineal massage, perineal lacerations, or episiotomy rates | Cochrane ROBT | High risk | High attrition rate, self-assessment by unblinded participants, unclear randomization |

[a]NIH QAT: NIH Quality Assessment Tool.

[b]OR: odds ratio.

[c]NICU: neonatal intensive care unit.

[d]IG: intervention group.

[e]CG: control group.

[f]ROBT: risk of bias tool.

### Risk of Bias in Included Studies

Bias ratings for all studies are included above in Tables 1-8. In total, 5 studies received a low risk rating, 11 studies received a fair risk rating, and 12 received a high risk rating. Reasons for the ratings were varied, and several studies had multiple reasons for increased risk of bias. Most commonly, studies with fair or high risk scores had issues with blinding (10 studies), high attrition (9 studies), or randomization (7 studies). Blinding issues most commonly revolved around patients and/or providers knowing a patient's allocation during the study. Randomization issues were varied and included unclear randomization schemes and lack of true randomization (being allocated by study staff). Other less common issues included low sample size (5 studies) and high rates of participant-reported outcomes (5 studies).

## Discussion

### Principal Findings

The findings from this systematic review suggest that available stand-alone mobile phone interventions show some positive changes in behavior and health outcomes in pregnant patients. Although findings were limited and some studies had high risk of bias, these early data suggest such interventions may have some ability to improve behaviors and health outcomes.

With regard to gestational weight gain, obesity, and physical activity, certain interventions did correlate with better health outcomes. In particular, there was less overall gestational weight gain in intervention users over the study period, but all three interventions used a multimodal intervention–either an app or text message in combination with another method of communication, such as social media or email. Regarding smoking cessation, controlled trials did not appear to exhibit an effect of interventions on improved rates of smoking. Regarding influenza vaccination, text message interventions did not improve rates of influenza vaccination. Finally, in the fourth group, which reviewed general prenatal health, interventions were associated with greater knowledge and positive health beliefs, but not with important health outcomes including cesarean delivery or other birth outcomes. Most studies we evaluated exhibited bias, most commonly with unclear blinding, high attrition and poor randomization, though also with low sample sizes, and self-reported outcomes, sometimes from unblinded participants.

Previous research has also explored the multimodal approach in this arena and found it effective. A Dutch study reviewed a 6-month mixed intervention involving Web-based, email, and provider-input components used with 1878 participants who were pregnant or contemplating pregnancy. The usability of the complete program was judged as *positive* or *very positive* by 54.7% of participants and study compliance was 64.86% (1218/1878) among all participants who activated the program. It was also associated with numerous improvements in nutrition and lifestyle behaviors such as improvement in adequate vegetable intake (26.3%, 95% CI 23.0-29.9), fruit intake (38.4%, 95% CI 34.5-42.5), folic acid use (56.3%, 95% CI 48.8-63.6), tobacco abstinence (35.1%, 95% CI 29.1-41.6), and alcohol abstinence (41.9%, 95% CI 35.2-48.9). The strongest effectiveness was for participating couples, which again may point to the multifactorial and social nature of successful interventions [57].

Chronic diseases continue to affect women at high rates during pregnancy and are also associated with increased risk of morbidity later in life, underscoring the need for continued exploration of efficient, wide-reaching interventions for monitoring and behavior modification. A very recent study with a focus on cardiovascular risk in pregnancy and the postpartum period emphasized that women with adverse obstetric outcomes such as preeclampsia, gestational hypertension, and gestational diabetes are at increased risk of developing CVD later in life. Specifically, women with preeclampsia have higher rates of overall CVD (relative risk [RR] 2.15; 95% CI 1.76-2.61), ischemic heart disease (RR 2.06; 95% CI 1.68-2.52), diabetes (RR 2.27; 95% CI 1.55-3.32), and premature cardiovascular death (RR 1.49; 95% CI 1.05-2.14) compared with women with uncomplicated pregnancies. The study authors urge that postpartum women with risk factors should be followed up vigilantly for blood pressure, BMI, waist circumference, lipid profile, fasting glucose, and oral glucose tolerance testing [58]. We propose monitoring and feedback for these patients be included in future mHealth interventions. Further exploration of wearable technologies including smartwatches, fitness bands, and other novel devices may assist with this endeavor.

Further research in this area could take multiple forms. For example, medication adherence for patients with diabetes and hypertension could be explored; a study of adolescents and young adults with sickle cell disease found that of proposed mobile phone app features for improving adherence, daily medication reminders were ranked first most frequently; this sentiment may be shared by pregnant patients in the same age group [59]. Further economic data may also be beneficial. One study of various mHealth interventions with reported economic evaluations found that 74.3% of interventions were cost-effective, economically beneficial, or cost saving [60]. This was briefly noted in one of our reviewed studies [36], but additional data on the topic are necessary. In addition, future work may investigate cross-platform technologies, such as those that are both stand-alone mobile phone platforms and available via the Web.

### Strengths and Limitations

The primary strength of this review is its inclusion of a wide variety of studies that investigated changes in behavior and clinical outcomes, both critical pieces necessary to evaluate novel behavioral technologies. Our study was focused on stand-alone interventions that did not require intensive clinical support and also reviewed data from high-income countries, which may provide more specifically applicable data for patients and physicians.

However, limitations include insufficiency of existing data and lack of granular clinical outcomes data in existing reports. Many included studies also exhibited high levels of bias with an unclear effect on results. In addition, no systematic evaluation of the interventions was performed (for example, using a specified taxonomy), which would allow more formal organization of intervention features themselves. Finally, intervention designs varied widely as did the measured outcomes and time frames of studies. All of these factors precluded the completion of a meaningful meta-analysis.

### Comparison With Existing Literature

A systematic review recently published in April 2018 by Overdijkink et al [61] described a similar review of studies employing text messages and mobile phone apps in pregnancy. Although there was an overlap in included studies, their methodology differed most notably because of the inclusion of telemedicine-based approaches. Most notably, at least five of their studies addressed gestational diabetes telemedicine and remote monitoring systems in which glucometers were coupled with mobile phone communications for nurse or physician feedback. In contrast, we aimed to find studies with minimal clinician input, preferring automated systems, and self-tracking technologies that supported or enhanced behavior changes without added clinician burdens. Furthermore, several of their included studies utilized primarily email and Web-based approaches, whereas we aimed to limit our review to mobile phone app and text-based technologies that could be implemented with use of phones or other primarily mobile technology. Despite these differences, we identified similarly that results are heterogeneous and that additional research is required to evaluate the effects of mHealth interventions on long-standing positive health outcomes.

### Conclusions

In high-income countries, utilization of mobile phone–based health behavior interventions in pregnancy demonstrates some correlation with positive beliefs, behaviors, and health outcomes. More effective interventions are multimodal in terms of features and tend to focus on healthy gestational weight gain. As mHealth interventions become increasingly available, future work must aim to maximize the clinical effectiveness of such interventions. As researchers, we should aim to broaden the scope of effective and sustainable interventions and continue to augment our care with appropriate evidence-based technologies.

## Acknowledgments

LY is supported by NICHD K12 HD050121-11. The funding source was not involved in study design; collection, analysis, and interpretation of data; writing of the report; and in the decision to submit the article for publication.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File , 19 KB - mhealth_v8i5e15111_app1.docx ]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analysis checklist.
[PDF File (Adobe PDF File), 524 KB - mhealth_v8i5e15111_app2.pdf ]

## References

1. Kaaja RJ, Greer IA. Manifestations of chronic disease during pregnancy. J Am Med Assoc 2005 Dec 7;294(21):2751-2757. [doi: 10.1001/jama.294.21.2751] [Medline: 16333011]
2. Walsh JM, McGowan CA, Mahony RM, Foley ME, McAuliffe FM. Obstetric and metabolic implications of excessive gestational weight gain in pregnancy. Obesity (Silver Spring) 2014 Jul;22(7):1594-1600 [FREE Full text] [doi: 10.1002/oby.20753] [Medline: 24715415]
3. Johansson AL, Dickman PW, Kramer MS, Cnattingius S. Maternal smoking and infant mortality: does quitting smoking reduce the risk of infant death? Epidemiology 2009 Jul;20(4):590-597. [doi: 10.1097/EDE.0b013e31819dcc6a] [Medline: 19289964]
4. Harrison AL, Shields N, Taylor NF, Frawley HC. Exercise improves glycaemic control in women diagnosed with gestational diabetes mellitus: a systematic review. J Physiother 2016 Oct;62(4):188-196 [FREE Full text] [doi: 10.1016/j.jphys.2016.08.003] [Medline: 27637772]

5. Centers for Disease Control and Prevention. Pregnancy Mortality Surveillance System URL: https://www.cdc.gov/reproductivehealth/maternal-mortality/pregnancy-mortality-surveillance-system.htm [accessed 2019-05-29]
6. Sieverdes JC, Treiber F, Jenkins C, Hermayer K. Improving diabetes management with mobile health technology. Am J Med Sci 2013 Apr;345(4):289-295. [doi: 10.1097/MAJ.0b013e3182896cee] [Medline: 23531961]
7. Scott-Sheldon LA, Lantini R, Jennings EG, Thind H, Rosen RK, Salmoirago-Blotcher E, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49 [FREE Full text] [doi: 10.2196/mhealth.5436] [Medline: 27207211]
8. 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]
9. Badawy SM, Barrera L, Sinno MG, Kaviany S, O'Dwyer LC, Kuhns LM. Text messaging and mobile phone apps as interventions to improve adherence in adolescents with chronic health conditions: a systematic review. JMIR Mhealth Uhealth 2017 May 15;5(5):e66 [FREE Full text] [doi: 10.2196/mhealth.7798] [Medline: 28506955]
10. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
11. Poushter J, Bishop C, Chwe H. Pew Research Center. 2018 Jun 19. Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones URL: https://www.pewresearch.org/global/2018/06/19/social-media-use-continues-to-rise-in-developing-countries-but-plateaus-across-developed-ones/ [accessed 2019-05-29]
12. Silver L. Pew Research Center. 2019 Feb 5. Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ [accessed 2019-05-29]
13. Urrutia RP, Berger AA, Ivins AA, Beckham AJ, Thorp JM, Nicholson WK. Internet use and access among pregnant women via computer and mobile phone: implications for delivery of perinatal care. JMIR Mhealth Uhealth 2015 Mar 30;3(1):e25 [FREE Full text] [doi: 10.2196/mhealth.3347] [Medline: 25835744]
14. Osma J, Barrera AZ, Ramphos E. Are pregnant and postpartum women interested in health-related apps? Implications for the prevention of perinatal depression. Cyberpsychol Behav Soc Netw 2016 Jun;19(6):412-415. [doi: 10.1089/cyber.2015.0549] [Medline: 27327069]
15. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 9;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
16. Mastrogiannis DS, Igwe E, Homko CJ. The role of telemedicine in the management of the pregnancy complicated by diabetes. Curr Diab Rep 2013 Feb;13(1):1-5. [doi: 10.1007/s11892-012-0352-x] [Medline: 23242646]
17. Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim K. Telemedicine interventions for gestational diabetes mellitus: A systematic review and meta-analysis. Diabetes Res Clin Pract 2015 Oct;110(1):1-9. [doi: 10.1016/j.diabres.2015.07.007] [Medline: 26264410]
18. Odibo IN, Wendel PJ, Magann EF. Telemedicine in obstetrics. Clin Obstet Gynecol 2013 Sep;56(3):422-433. [doi: 10.1097/GRF.0b013e318290fef0] [Medline: 23903374]
19. Ahl SG, Jovanovič L, Hod M. Technology and pregnancy. Diabetes Technol Ther 2013 Feb;15(Suppl 1):S75-S82. [doi: 10.1089/dia.2013.1508] [Medline: 23441710]
20. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. PLoS One 2016;11(5):e0154664 [FREE Full text] [doi: 10.1371/journal.pone.0154664] [Medline: 27144393]
21. Coleman J, Bohlin KC, Thorson A, Black V, Mechael P, Mangxaba J, et al. Effectiveness of an SMS-based maternal mHealth intervention to improve clinical outcomes of HIV-positive pregnant women. AIDS Care 2017 Jul;29(7):890-897. [doi: 10.1080/09540121.2017.1280126] [Medline: 28107805]
22. Joos O, Silva R, Amouzou A, Moulton LH, Perin J, Bryce J, et al. Evaluation of a mHealth data quality intervention to improve documentation of pregnancy outcomes by health surveillance assistants in Malawi: a cluster randomized trial. PLoS One 2016;11(1):e0145238 [FREE Full text] [doi: 10.1371/journal.pone.0145238] [Medline: 26731401]
23. Hussain T, Smith P, Yee LM. CRD Database - University of York. 2016. A Systematic Review of Smartphone-based mHealth Interventions in Pregnancy on Maternal Behaviors and Maternal and Fetal Health Outcomes in Developed Countries URL: http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42016051055 [accessed 2018-12-31]
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
25. United Nations. World Economic Situation and Prospects 2012. New York: United Nations; 2012.
26. abstrackr. URL: http://abstrackr.cebm.brown.edu/ [accessed 2020-02-25]
27. Rayyan QCRI. URL: https://rayyan.qcri.org [accessed 2020-02-25]

28. Soltani H, Duxbury A, Arden M, Dearden A, Furness P, Garland C. Maternal obesity management using mobile technology: a feasibility study to evaluate a text messaging based complex intervention during pregnancy. J Obes 2015;2015:814830 [FREE Full text] [doi: 10.1155/2015/814830] [Medline: 25960889]
29. Higgins JP, Altman DG, Sterne JA. Cochrane Handbook for Systematic Reviews of Interventions. Chapter 8: Assessing Risk of Bias in Included Studies URL: https://handbook-5-1.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm [accessed 2018-12-20] [WebCite Cache ID 74oFdbnqA]
30. Choi J, Lee JH, Vittinghoff E, Fukuoka Y. mHealth physical activity intervention: a randomized pilot study in physically inactive pregnant women. Matern Child Health J 2016 May;20(5):1091-1101 [FREE Full text] [doi: 10.1007/s10995-015-1895-7] [Medline: 26649879]
31. Herring SJ, Cruice JF, Bennett GG, Rose MZ, Davey A, Foster GD. Preventing excessive gestational weight gain among African American women: a randomized clinical trial. Obesity (Silver Spring) 2016 Jan;24(1):30-36 [FREE Full text] [doi: 10.1002/oby.21240] [Medline: 26592857]
32. Dodd JM, Louise J, Cramp C, Grivell RM, Moran LJ, Deussen AR. Evaluation of a smartphone nutrition and physical activity application to provide lifestyle advice to pregnant women: The SNAPP randomised trial. Matern Child Nutr 2018 Jan;14(1). [doi: 10.1111/mcn.12502] [Medline: 28836373]
33. Willcox JC, Wilkinson SA, Lappas M, Ball K, Crawford D, McCarthy E, et al. A mobile health intervention promoting healthy gestational weight gain for women entering pregnancy at a high body mass index: the txt4two pilot randomised controlled trial. Br J Obstet Gynaecol 2017 Oct;124(11):1718-1728 [FREE Full text] [doi: 10.1111/1471-0528.14552] [Medline: 28220604]
34. Pollak KI, Alexander SC, Bennett G, Lyna P, Coffman CJ, Bilheimer A, et al. Weight-related SMS texts promoting appropriate pregnancy weight gain: a pilot study. Patient Educ Couns 2014 Nov;97(2):256-260 [FREE Full text] [doi: 10.1016/j.pec.2014.07.030] [Medline: 25153313]
35. Huberty JL, Buman MP, Leiferman JA, Bushar J, Hekler EB, Adams MA. Dose and timing of text messages for increasing physical activity among pregnant women: a randomized controlled trial. Transl Behav Med 2017 Jun;7(2):212-223 [FREE Full text] [doi: 10.1007/s13142-016-0445-1] [Medline: 27800565]
36. Redman LM, Gilmore LA, Breaux J, Thomas DM, Elkind-Hirsch K, Stewart T, et al. Effectiveness of SmartMoms, a novel eHealth intervention for management of gestational weight gain: randomized controlled pilot trial. JMIR Mhealth Uhealth 2017 Sep 13;5(9):e133 [FREE Full text] [doi: 10.2196/mhealth.8228] [Medline: 28903892]
37. Kennelly MA, Ainscough K, Lindsay KL, O Sullivan E, Gibney ER, McCarthy M, et al. Pregnancy exercise and nutrition with smartphone application support: a randomized controlled trial. Obstet Gynecol 2018 May;131(5):818-826. [doi: 10.1097/AOG.0000000000002582] [Medline: 29630009]
38. Abroms LC, Johnson PR, Heminger CL, van Alstyne JM, Leavitt LE, Schindler-Ruwisch JM, et al. Quit4baby: results from a pilot test of a mobile smoking cessation program for pregnant women. JMIR Mhealth Uhealth 2015 Jan 23;3(1):e10 [FREE Full text] [doi: 10.2196/mhealth.3846] [Medline: 25650765]
39. Fujioka N, Kobayashi T, Turale S. Short-term behavioral changes in pregnant women after a quit-smoking program via e-learning: a descriptive study from Japan. Nurs Health Sci 2012 Sep;14(3):304-311. [doi: 10.1111/j.1442-2018.2012.00702.x] [Medline: 22950611]
40. Abroms LC, Johnson PR, Leavitt LE, Cleary SD, Bushar J, Brandon TH, et al. A randomized trial of text messaging for smoking cessation in pregnant women. Am J Prev Med 2017 Dec;53(6):781-790 [FREE Full text] [doi: 10.1016/j.amepre.2017.08.002] [Medline: 28982527]
41. Abroms LC, Chiang S, Macherelli L, Leavitt L, Montgomery M. Assessing the national cancer institute's SmokefreeMOM text-messaging program for pregnant smokers: pilot randomized trial. J Med Internet Res 2017 Oct 3;19(10):e333 [FREE Full text] [doi: 10.2196/jmir.8411] [Medline: 28974483]
42. Naughton F, Prevost A, Gilbert H, Sutton S. Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). Nicotine Tob Res 2012 May;14(5):569-577. [doi: 10.1093/ntr/ntr254] [Medline: 22311960]
43. Pollak KI, Lyna P, Bilheimer A, Farrell D, Gao X, Swamy GK, et al. A pilot study testing SMS text delivered scheduled gradual reduction to pregnant smokers. Nicotine Tob Res 2013 Oct;15(10):1773-1776 [FREE Full text] [doi: 10.1093/ntr/ntt045] [Medline: 23569007]
44. Tombor I, Beard E, Brown J, Shahab L, Michie S, West R. Randomized factorial experiment of components of the SmokeFree Baby smartphone application to aid smoking cessation in pregnancy. Transl Behav Med 2019 Jul 16;9(4):583-593 [FREE Full text] [doi: 10.1093/tbm/iby073] [Medline: 30011020]
45. Forinash AB, Yancey A, Chamness D, Koerner J, Inteso C, Miller C, et al. Smoking cessation following text message intervention in pregnant women. Ann Pharmacother 2018 Nov;52(11):1109-1116. [doi: 10.1177/1060028018780448] [Medline: 29857773]
46. Naughton F, Cooper S, Foster K, Emery J, Leonardi-Bee J, Sutton S, et al. Large multi-centre pilot randomized controlled trial testing a low-cost, tailored, self-help smoking cessation text message intervention for pregnant smokers (MiQuit). Addiction 2017 Jul;112(7):1238-1249 [FREE Full text] [doi: 10.1111/add.13802] [Medline: 28239919]

47. Moniz MH, Hasley S, Meyn LA, Beigi RH. Improving influenza vaccination rates in pregnancy through text messaging: a randomized controlled trial. Obstet Gynecol 2013 Apr;121(4):734-740. [doi: 10.1097/AOG.0b013e31828642b1] [Medline: 23635672]
48. Yudin MH, Mistry N, de Souza LR, Besel K, Patel V, Mejia SB, et al. Text messages for influenza vaccination among pregnant women: A randomized controlled trial. Vaccine 2017 Feb 1;35(5):842-848. [doi: 10.1016/j.vaccine.2016.12.002] [Medline: 28062124]
49. Moniz M, Meyn L, Beigi R. Text messaging to improve preventive health attitudes and behaviors during pregnancy: a prospective cohort analysis. J Reprod Med 2015;60(9-10):378-382. [Medline: 26592061]
50. Dalrymple PW, Rogers M, Zach L, Turner K, Green M. Collaborating to develop and test an enhanced text messaging system to encourage health information seeking. J Med Libr Assoc 2013 Jul;101(3):224-227 [FREE Full text] [doi: 10.3163/1536-5050.101.3.014] [Medline: 23930095]
51. Bush J, Barlow DE, Echols J, Wilkerson J, Bellevin K. Impact of a mobile health application on user engagement and pregnancy outcomes among Wyoming medicaid members. Telemed J E Health 2017 Nov;23(11):891-898 [FREE Full text] [doi: 10.1089/tmj.2016.0242] [Medline: 28481167]
52. Krishnamurti T, Davis AL, Wong-Parodi G, Fischhoff B, Sadovsky Y, Simhan HN. Development and testing of the MyHealthyPregnancy app: a behavioral decision research-based tool for assessing and communicating pregnancy risk. JMIR Mhealth Uhealth 2017 Apr 10;5(4):e42 [FREE Full text] [doi: 10.2196/mhealth.7036] [Medline: 28396302]
53. Ledford CJ, Canzona MR, Cafferty LA, Hodge JA. Mobile application as a prenatal education and engagement tool: a randomized controlled pilot. Patient Educ Couns 2016 Apr;99(4):578-582. [doi: 10.1016/j.pec.2015.11.006] [Medline: 26610389]
54. Evans WD, Bihm JW, 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]
55. 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]
56. Takeuchi S, Horiuchi S. Randomised controlled trial using smartphone website vs leaflet to support antenatal perineal massage practice for pregnant women. Women Birth 2016 Oct;29(5):430-435 [FREE Full text] [doi: 10.1016/j.wombi.2016.01.010] [Medline: 26906970]
57. van Dijk MR, Huijgen NA, Willemsen SP, Laven JS, Steegers EA, Steegers-Theunissen RP. Impact of an mHealth platform for pregnancy on nutrition and lifestyle of the reproductive population: a survey. JMIR Mhealth Uhealth 2016 May 27;4(2):e53 [FREE Full text] [doi: 10.2196/mhealth.5197] [Medline: 27234926]
58. Smith GN, Louis JM, Saade GR. Pregnancy and the postpartum period as an opportunity for cardiovascular risk identification and management. Obstet Gynecol 2019 Oct;134(4):851-862. [doi: 10.1097/AOG.0000000000003363] [Medline: 31503139]
59. Badawy SM, Thompson AA, Liem RI. Technology access and smartphone app preferences for medication adherence in adolescents and young adults with sickle cell disease. Pediatr Blood Cancer 2016 May;63(5):848-852. [doi: 10.1002/pbc.25905] [Medline: 26844685]
60. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
61. Overdijkink SB, Velu AV, Rosman AN, van Beukering MD, Kok M, Steegers-Theunissen RP. The usability and effectiveness of mobile health technology-based lifestyle and medical intervention apps supporting health care during pregnancy: systematic review. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e109 [FREE Full text] [doi: 10.2196/mhealth.8834] [Medline: 29691216]


## Abbreviations

**CVD:** cardiovascular disease
**mHealth:** mobile health
**NIT:** National Institutes of Health
**NIH QAT:** NIH Quality Assessment Tool
**PRISMA:** Preferred Reporting Items for Systematic Reviews and Meta-Analysis
**RCT:** randomized controlled trial
**RR:** relative risk

*Edited by G Eysenbach; submitted 26.06.19; peer-reviewed by J Coleman, S Badawy, MDG Pimentel, SGS Shah; comments to author 03.10.19; revised version received 22.11.19; accepted 26.01.20; published 28.05.20.*

*Please cite as:*
*Hussain T, Smith P, Yee LM*
*Mobile Phone–Based Behavioral Interventions in Pregnancy to Promote Maternal and Fetal Health in High-Income Countries: Systematic Review*
*JMIR Mhealth Uhealth 2020;8(5):e15111*
*URL: https://mhealth.jmir.org/2020/5/e15111*
*doi:10.2196/15111*
*PMID:32463373*

Original Paper

# Mobile App to Improve House Officers' Adherence to Advanced Cardiac Life Support Guidelines: Quality Improvement Study

Vittal Hejjaji[1], MD; Ali O Malik[1], MD; Poghni A Peri-Okonny[1], MD, MSc; Merrill Thomas[1], MD; Yuanyuan Tang[1], PhD; David Wooldridge[2], MD; John A Spertus[1], MD, MPH; Paul S Chan[1], MD, MSc

[1]Department of Cardiovascular Medicine, Saint Luke's Mid America Heart Institute, University of Missouri Kansas City, Kansas City, MO, United States

[2]Department of Internal Medicine, University of Missouri Kansas City, Kansas City, MO, United States

**Corresponding Author:**
Vittal Hejjaji, MD
Department of Cardiovascular Medicine
Saint Luke's Mid America Heart Institute
University of Missouri Kansas City
4401 Wornall Rd
Kansas City, MO
United States
Phone: 1 2162781685
Email: hejjajiv@umkc.edu

## Abstract

**Background:** Effective and timely delivery of cardiac arrest interventions during in-hospital cardiac arrest resuscitation is associated with greater survival. Whether a mobile app that provides timely reminders of critical interventions improves adherence to Advanced Cardiovascular Life Support (ACLS) guidelines among house officers, who may lack experience in leading resuscitations, remains unknown.

**Objective:** The aim of this study was to assess the impact of a commercially available, dynamic mobile app on house officers' adherence to ACLS guidelines.

**Methods:** As part of a quality improvement initiative, internal medicine house officers were invited to participate and randomized to lead 2 consecutive cardiac arrest simulations, one with a novel mobile app and one without a novel mobile app. All simulations included 4 cycles of cardiopulmonary resuscitation with different cardiac arrest rhythms and were video recorded. The coprimary end points were chest compression fraction and number of correct interventions in each simulation. The secondary end point was incorrect interventions, defined as interventions not indicated by the 2015 ACLS guidelines. Paired *t* tests compared performance with and without the mobile app.

**Results:** Among 53 house officers, 26 house officers were randomized to lead the first simulation with the mobile app, and 27 house officers were randomized to do so without the app. Use of the mobile app was associated with a higher number of correct ACLS interventions (out of 7; mean 6.2 vs 5.1; absolute difference 1.1 [95% CI 0.6 to 1.6]; $P<.001$) as well as fewer incorrect ACLS interventions (mean 0.3 vs 1.0; absolute difference –0.7 [95% CI –0.3 to –1.0]; $P<.001$). Simulations with the mobile app also had a marginally higher chest compression fraction (mean 90.9% vs 89.0%; absolute difference 1.9% [95% CI 0.6% to 3.4%]; $P=.007$).

**Conclusions:** This proof-of-concept study suggests that this novel mobile app may improve adherence to ACLS protocols, but its effectiveness on survival in real-world resuscitations remains unknown.

***(JMIR Mhealth Uhealth 2020;8(5):e15762)*** doi:10.2196/15762

**KEYWORDS**

cardiac arrest; advanced cardiac life support; mHealth; quality improvement; medical education

## Introduction

### Background

Advanced Cardiovascular Life Support (ACLS) guidelines provide evidence-based algorithms to optimize the likelihood of survival in patients with in-hospital cardiac arrest (IHCA) [1]. Several ACLS components as well as the recent 2015 American Heart Association (AHA) resuscitation guidelines are critical for high-quality cardiopulmonary resuscitation (CPR) [2]. These include minimizing interruptions to chest compressions to achieve a high compression fraction, timely defibrillation, accurate and timely dosing of vasoactive drugs, and avoiding inappropriate ACLS treatments (eg, atropine for asystole). However, there is wide hospital-level variability in the delivery of timely defibrillation and epinephrine [3,4]. Whether a novel and portable mobile app that tracks critical interventions and provides real-time reminders and dosing guidance during a resuscitation improves adherence to ACLS guidelines remains unclear, but this would be important to understand as it may help reduce variability in the delivery of potentially lifesaving interventions.

Acute resuscitations are often chaotic, which may contribute to the variability in adherence to ACLS guidelines, thus attenuating the benefits of ACLS [5]. At many hospitals, this is compounded by the fact that house officers often lead resuscitations, although they may have little experience and feel inadequately prepared [6]. Use of an auxiliary mobile phone tool that includes real-time prompts for when to defibrillate or administer vasoactive medications, separate clocks for tracking chest compression duration and time since last defibrillation or vasoactive medication intervention, and dosing guidance for vasoactive medications (such as for epinephrine and amiodarone) may be an important adjunct to facilitate high-quality CPR among house officers. Moreover, such a tool has the potential to minimize interventions that are no longer recommended or even inappropriate per the current ACLS guidelines, such as procainamide or sodium bicarbonate.

### Objective

As part of a quality improvement program, we conducted a clinical trial wherein each house officer was randomized to perform a cardiac arrest simulation with and without a mobile app designed to support the use of ACLS guidelines. We examined whether use of the mobile app was associated with a higher number of correct ACLS interventions performed and chest compression fraction, as well as fewer inappropriate interventions, in a simulation setting.

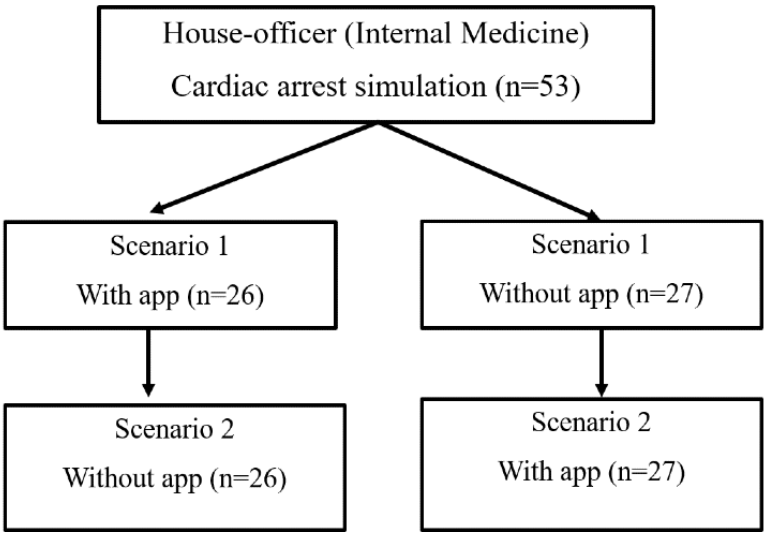
## Methods

### Study Population and Setting

This study required all participants to perform 2 resuscitation simulations, one with and one without the ACLS mobile app. The study was a quality improvement initiative within the internal medicine residency program at the University of Missouri-Kansas City School of Medicine. All categorical and preliminary house officers were invited to participate by the residency director (DW). Participation was voluntary. A total of 53 ACLS-certified house officers out of 60 participated in the study between August 6, 2018, and September 7, 2018. The study was approved by the institutional review board at Saint Luke's Hospital in Kansas City, Missouri (approved on November 4, 2018), as a quality improvement study. Regardless, individual informed consent was obtained for randomization and study participation.

Study simulations were conducted at an AHA-certified training center in Kansas City, Missouri, which frequently conducts mock codes to mimic a real-world inpatient setting. All house officer participants completed 2 different 10- to 12-min simulations consecutively—one with the mobile app and the other without the mobile app. Each of the 2 simulations comprised 4 cycles of CPR, rhythm identification, and defibrillation or drug administration, each cycle lasting for 2 min. A random number generator determined whether the house officer performed the first or second simulation with the mobile app tool (Figure 1).

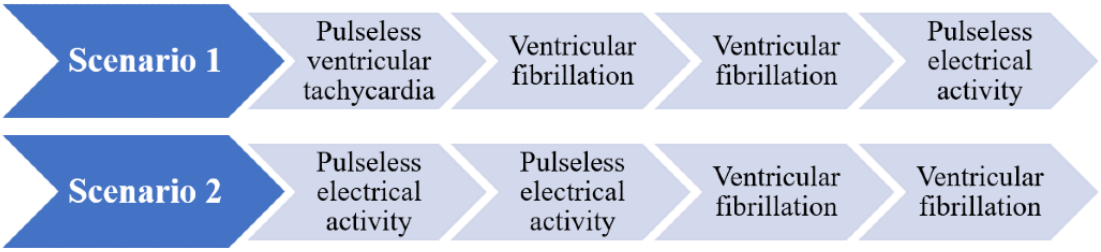
**Figure 1.** Scheme of randomization for leading each scenario with or without the mobile app.



The different cardiac arrest rhythms used in each scenario are shown in Figure 2. Although the scenario in which the house officer would use the mobile app was decided randomly, the sequence of the 2 simulations itself varied depending on the day of the week on which the house officer participated in the study, such that half the days started with scenario 1 and the other half started with scenario 2.

All resuscitation equipment, such as the defibrillator, medications, intravenous lines, and airway, were obtained from a real-world crash cart. Two certified training personnel with predefined roles were assigned to each simulation. For data acquisition, both simulations (with and without the mobile app) were videotaped to ensure accurate documentation and timing of interventions for end point assessment.

**Figure 2.** Scheme of cardiac arrest rhythms used in each scenario.




## Study Intervention

The intervention of interest was a mobile app for ACLS delivery, which can be downloaded onto any smartphone and is compatible with both Android and iPhone operating system platforms. The Redivus Code Blue app was created and licensed by Redivus Health, and it is commercially available on Google Play and Apple App Store.

The goal of the mobile app is to offload team leader burden and simultaneously reduce delays in ACLS treatments, while increasing adherence to ACLS guidelines and AHA quality metrics. Activation of the *Code Blue* interface within the app initiates 2 sets of timers. One timer is for the CPR cycle and provides prompts for regular pulse checks at 2-min intervals, as recommended in ACLS. The second timer provides prompts for when the next defibrillation attempt or vasoactive medication dose should be administered, depending on the identified cardiac arrest rhythm (Multimedia Appendices 1 and 2 show the screenshots of the mobile app depicting the sequence of events). These time intervals are based on existing resuscitation guidelines (eg, 3 to 5 min between subsequent epinephrine doses). In addition to avoiding delays in administering treatments, the mobile app also provides guidance for the next recommended intervention and dosing in the ACLS algorithm to prevent deviation from guideline recommendations.

For this study, once informed consent was obtained, each participant was oriented to the mobile app through a 10-min prerecorded video, as well as to the simulation environment including personnel. All participants were permitted to use other decision aids (eg, ACLS algorithm card) they would normally use in either simulation to reflect usual care. They were then randomized to perform the first or second simulation with the mobile app.

## Study Outcomes

Our coprimary outcomes were overall compression fraction and the number of correct ACLS interventions. For the outcome of correct ACLS interventions, each of the 2 scenarios had a total of 7 correct interventions (defibrillation, epinephrine, and amiodarone). A correct intervention was defined as timely administration of the correct intervention (defibrillation or medication, including drug dose) per the ACLS protocol. If a correct intervention was skipped or the wrong medication dose was administered, a point was deducted from the total score of 7. In addition, we examined as a secondary outcome, the number of incorrect interventions, defined as an inappropriate ACLS treatment based on cardiac arrest rhythm. Examples of incorrect treatment include defibrillation for a nonshockable cardiac arrest rhythm and medications not recommended in the current ACLS guidelines (eg, atropine, bicarbonate, and procainamide).

## Statistical Analysis

Each simulation recording was viewed and graded by research personnel, equally divided among the 4 individuals, with 10% of the simulations validated by a second reviewer, who separately graded the simulation to ensure accuracy and consistency in scoring. Discordant ratings between 2 graders were resolved by a third researcher (PC). Each research personnel grading the simulations was provided with a structured score sheet and instructions detailing the use of the score sheet. Compression fraction was defined as the total amount of time (in seconds) taken for performing CPR divided by the total time of the simulation, and this was obtained from the time stamp for each period in the simulation scenario when CPR was and was not being administered. Similarly, for the number of correct ACLS interventions, we used the time stamp to ensure timely delivery of medications and defibrillations. For epinephrine and amiodarone dosing, where the recommended interval is 3 to 5 min between doses, we allowed a 15-second grace window on either side of this time interval and graded all medication doses ordered between 2:45 and 5:15 min as correct, as long as the medication dosing was also correct (1 mg of epinephrine; 300 mg and 150 mg for the first and second doses of amiodarone, respectively) and as long as it was the appropriate treatment for the identified rhythm.

For the coprimary outcomes, we used a paired *t* test to assess differences. For the secondary outcome of incorrect number of

interventions, we used the Wilcoxon signed-rank test to estimate the effect of the mobile app on ACLS adherence. In addition, we conducted additional interaction analyses to assess whether the study outcome results differed by the house officer's level of training (first vs second vs third year), whether the mobile app was used for the first or the second simulation to rule out a "learning effect," and whether the house officer had previous experience with leading a real-world resuscitation event in the hospital. To accomplish this, we constructed 3 multivariable models with correct and incorrect interventions as the outcome and separately evaluated interaction terms between use of the mobile app and these 3 study factors, using linear regression. All analyses were performed using SAS version 9.4 (SAS institute, Cary, North Carolina, United States) and were evaluated at a two-sided significance level of .05.

# Results

## User Statistics

Of the 53 participants in the study, 24 (45%) participants were first-year, 15 (28%) participants were second-year, and 14 (26%) participants were third-year house officers. Approximately half of the (26/53, 49%) house officers were randomized to use the mobile app for their first simulation. A total of 10 house officers (10/53, 19%) had previously led at least one real-world resuscitation for cardiac arrest in the hospital.

## Evaluation Outcomes

Overall, use of the mobile app resulted in a small improvement in compression fraction (mean 90.9% vs 89.0%; absolute difference 1.9% [95% CI 0.6% to 3.4%]; *P*=.007; Table 1).

**Table 1.** Effect of the mobile app on study outcomes.

| Outcome variable | Without the app, mean (SD) | With the app, mean (SD) | Absolute difference with the app, (95% CI) | *P* value |
|---|---|---|---|---|
| Compression fraction | 89.0% (5.0%) | 90.9% (2.3%) | 1.9% (0.6% to 3.4%) | .007 |
| Number of correct interventions | 5.1 (1.6) | 6.2 (1.1) | 1.1 (0.6 to 1.6) | <.001 |
| Number of incorrect interventions | 1.0 (1.3) | 0.3 (0.6) | −0.7 (−0.3 to −1.0) | <.001 |

The number of correct ACLS interventions (out of 7) was significantly higher among simulations in which the mobile app was used (mean 6.2 vs 5.1; absolute difference 1.1 [95% CI 0.6 to 1.6]; *P*<.001). The reasons for not receiving credit for a correct intervention during the simulations, stratified by mobile app use, are outlined in Table 2.

**Table 2.** Most common reasons for not performing a correct Advanced Cardiovascular Life Support intervention.

| Reason for missing intervention | Total, N | Without the app, n | With the app, n |
|---|---|---|---|
| **Epinephrine** | | | |
| **Incorrect epinephrine timing** | | | |
| First dose | 15 | 13 | 2 |
| Second dose | 27 | 20 | 7 |
| **Amiodarone** | | | |
| Failed to give second dose of amiodarone | 28 | 19 | 9 |
| Failed to give any dose of amiodarone | 17 | 13 | 4 |
| **Incorrect amiodarone dose** | | | |
| First dose | 8 | 6 | 2 |
| Second dose | 5 | 1 | 4 |
| Incorrect amiodarone timing | 10 | 5 | 5 |
| **Chest compressions** | | | |
| Pulse check at irregular intervals affecting overall compression fraction | 7 | 7 | 0 |

The most common reasons for missing a correct ACLS intervention were incorrect timing of epinephrine dose (n=42) and not administering amiodarone at all when indicated by ACLS guidelines (n=45). These occurred more commonly in simulations performed with usual care than with the mobile app.

For our secondary outcome, use of the mobile app also resulted in a fewer number of incorrect interventions (mean 0.3 vs 1.0; absolute difference −0.7 [95% CI −0.3 to −1.0]; *P*<.001). The most common reasons for an incorrect intervention are summarized in Table 3.

**Table 3.** Most common incorrect Advanced Cardiovascular Life Support interventions performed.

| Incorrect intervention | Total, N | Without the app, n | With the app, n |
|---|---|---|---|
| Incorrect rhythm identification | 17 | 10 | 7 |
| Incorrect administration of epinephrine and failure to defibrillate pulseless ventricular tachycardia/ventricular fibrillation | 13 | 10 | 3 |
| Inappropriate defibrillation for PEA[a] | 8 | 7 | 1 |
| Used atropine to treat PEA | 7 | 5 | 2 |
| Checked blood pressure during chest compressions | 6 | 3 | 3 |

[a]PEA: pulseless electrical activity.

The most common reasons included incorrect rhythm identification (n=17), incorrectly administering epinephrine and not performing defibrillation for a shockable cardiac arrest rhythm (n=13), inappropriate defibrillation for a nonshockable pulseless electrical activity (PEA) rhythm (n=8), and use of atropine for a PEA cardiac arrest rhythm (n=7). Use of the mobile app resulted in a numerically lower number of each of these incorrect interventions.

Finally, there were no significant interactions (all *P* values for interaction were >.31) between the use of the mobile app and the year of house officer training, whether the mobile app was used for the first or second simulation, and whether the participant had previously led a resuscitation in the hospital setting for either the end point of the number of correct interventions or the number of incorrect interventions (Table 4).

After the completion of the 2 simulations, participants were given the opportunity to describe their experience using the mobile app, and the most common comments reported by 2 or more participants are summarized in Textbox 1.

**Table 4.** Interaction analyses for the end points of total correct interventions and total incorrect interventions.

| Interaction variables | *P* value |
|---|---|
| **Number of correct interventions** | |
| App usage x house officer training level | >.99 |
| App usage x sequence[a] | .32 |
| App usage x previous experience in leading codes | .50 |
| **Number of incorrect interventions** | |
| App usage x house officer training level | .50 |
| App usage x sequence[a] | .31 |
| App usage x previous experience in leading codes | .52 |

[a]Indicates the sequence of simulations (whether mobile app was used with the first or second simulation).

**Textbox 1.** Qualitative data from the participants.

House officers' comments postintervention:

1. Advantages
   - "The app was most useful for timing of cycles and Epi intervals"
   - "Good experience. I like the functionality of the app. If the documents section can be integrated into the existing EMR so that the physician can look at comorbidities and history that would be very helpful."
   - "I could definitely see an improvement in adhering to the guidelines using the app. At our institution nursing staff keeps track of time. Having to keep time threw me off a little during the simulation without the app. Overall it was a great experience using the app"
   - "If you don't remember the sequence the app prompts you which is great. If you do remember the sequence the app still keeps all the timing appropriate and code running smoothly"
   - "Although this study was among resident doctors due to the fact that it is currently expected of all doctors to be well versed in running code blues, with apps like these it should be possible for any health care professional to successfully run a code blue"
2. Limitations/suggestions for improvement
   - "I would strongly suggest use vibratory cues for the timing. I found myself looking down too much at the app and not enough at assessing clinical status"
   - "My only concern with the app is for some medications, it doesn't specify the dose until it is time for it to be given. It would be nice if the amount of the next dose was displayed, so you could ready it in advance"

## Discussion

### Principal Findings

In this quality improvement trial to improve the rates of ACLS adherence among internal medicine house officers in simulations of IHCAs, we evaluated the impact of a dynamic mobile app that provides timely reminders and dosing guidance for ACLS interventions to the resuscitation team leader. Using each participant as their own control, we found that the use of this mobile app increased overall compression fraction and the number of correct interventions. Moreover, the use of the mobile app resulted in 19 fewer occasions of incorrect interventions delivered. In addition, the results for all 3 outcomes were not different, regardless of the house officer's year of training, previous experience with leading in-hospital resuscitations, and whether the app was used for the first or second simulation. This study's findings suggest that the use of novel assistive technologies such as this mobile app could improve adherence to ACLS guidelines in hospitalized patients with cardiac arrest.

### Comparison With Previous Work

To the best of our knowledge, this is the first study in the United States to have evaluated the use of a cardiac arrest mobile app among house officers. Most teaching hospitals have resuscitation teams structured around house officers being the team leader [7]. The chaotic and often disorganized environment of an emergent cardiac arrest resuscitation, as well as house officers' infrequent exposure to leading resuscitations, makes house officers particularly suitable candidates for the use of mobile apps, such as the one used in this study, to minimize team leader burden. Dynamic decision aids that can reduce this burden on physician trainees have the potential to improve resuscitation care by increasing ACLS adherence and the overall quality of CPR delivered [8]. Improving adherence and standardizing care are important, as delays in the administration of defibrillation for a shockable rhythm and epinephrine for a nonshockable rhythm have been associated with worse survival outcomes for IHCA [9-11]. Furthermore, ACLS and the 2015 AHA resuscitation guidelines have highlighted the importance of effective chest compressions throughout the resuscitation process; therefore, minimizing interruptions to chest compressions should be a primary goal of all resuscitations [12].

Overall, this study found that house officers achieved a 22% relative increase in the number of correct ACLS interventions (6.2 vs 5.1). We found that the use of the mobile app ensured proper timing and dosing of the vasoactive medications, epinephrine and amiodarone, and failure to do so was the most common reason for missing a correct ACLS intervention in this study (see Table 2). Less robust results (15% relative increase) were observed in another study examining the use of a different electronic decision support tool among fourth-year medical students [8]. Moreover, this other study was conducted before the 2015 AHA/ACLS update, which delineated the importance of compression fraction and recommended an average fraction of more than 60% to be ideal [13]. Although this study showed a small improvement in compression fraction with the use of the mobile app, it remained above 60% and, more importantly, was not lower than that with usual care, as there may be concerns that decision tools may increase the frequency of interruptions to chest compressions [14].

Beyond improving adherence to correct ACLS interventions, we also found that it decreased the delivery of incorrect interventions. Inaccurate rhythm identification by house officers was the most common error and an issue for which the mobile app provides no benefit, as the app relies on the house officer to correctly identify the cardiac arrest rhythm to trigger the proper algorithm. We found that there were 10 inaccurate rhythm identifications in simulations without the mobile app and 7 inaccurate rhythm identifications with the app. The other reasons for an incorrect ACLS intervention were areas for which the mobile app can provide benefit. Administering epinephrine without defibrillating a shockable cardiac arrest rhythm or

defibrillating a nonshockable rhythm clearly indicates difficulty of the team leader to recall the correct ACLS algorithm, and this was more common in simulations without the mobile app than with its use (see Table 3). Given that the use of the mobile app significantly reduced the number of potentially life-threatening incorrect interventions, we believe it to be extremely useful in a clinical setting, but further studies with real-world use of the app will determine if the differences noticed in this analysis are clinically relevant.

### Limitations

This study should be interpreted in the context of several limitations. First, the mobile ACLS app was evaluated in a controlled simulation setting, and its use and impact during emergent hospital resuscitations were not evaluated. As such, this trial is a proof-of-concept study of the mobile app's value, but it will require further validation in real-world settings. This is particularly important because the potential benefits of using a mobile app are dependent on the app being used and the frequency and consistency of launching the app, and any delays in doing so are unknown and need to be examined in future studies. Moreover, the impact of the mobile app on overall survival rates for IHCA is unknown. Second, other components of resuscitation quality, such as ventilation and chest compression depth and rate, are not influenced by the mobile app. Therefore, the mobile app is only able to affect some, but not all, aspects of emergent resuscitation care. Third, this mobile app was evaluated among house officers only, and its benefits may not be generalizable to more experienced physicians. It is possible that experienced physicians have fewer gaps in the quality of their acute resuscitation care and that the mobile app would have little benefit. It is also possible that house officers attend more acute resuscitations in the hospital than other physicians and that the mobile app may show even greater benefit in physicians who have long completed their training. Fourth, this was a single-center study, and the findings may not reflect practice patterns at other institutions. Finally, the mobile app was designed to reduce team leader burden, but it has no direct effect on the team leader's leadership skills, which can play an important role in resuscitation care variability.

### Conclusions

The use of a novel mobile app by house officers during simulations for IHCA was associated with better adherence to ACLS performance measures, a lower rate of incorrect interventions, and a mildly higher chest compression fraction. Given these promising findings, further testing in real-world care should be conducted.

## Acknowledgments

The authors wish to acknowledge Audrey Lu for her assistance in the simulation center and the research project. VH, AM, and PO are supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number T32HL110837. PC is supported by an R01 grant (1R01HL123980) from the National Heart, Lung, and Blood Institute. The content herein is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The mobile app is licensed by Redivus Health, who provided it free of cost for the study. The cost of the simulation center and videotaping was also borne by Redivus Health. None of the employees of Redivus were involved in the data analysis or in the drafting or critical revisions of this paper. Moreover, none of the study authors received financial or other remuneration from the company for conducting the study.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Mobile app screenshots depicting the sequence of events for a nonshockable rhythm. A: The appropriate rhythm type is selected by the user. B: The app prompts the user to use epinephrine. C: Once epinephrine is delivered, a separate timer starts for the next dose of epinephrine. VFIB: ventricular fibrillation; VTACH: ventricular tachycardia; PEA: pulseless electrical activity; AED: automated external defibrillator; Defib: defibrillation.
[PNG File , 210 KB - mhealth_v8i5e15762_app1.png ]

Multimedia Appendix 2
Mobile app screenshots depicting the sequence of events for a shockable rhythm. A: The appropriate rhythm type is selected by the user. B: The app prompts the user to use the defibrillator by following steps 1 to 4, thereby avoiding chest compression interruptions during defibrillation. C: Once shock is delivered, a separate timer starts for the next dose of amiodarone. VFIB: ventricular fibrillation; VTACH: ventricular tachycardia; PEA: pulseless electrical activity.
[PNG File , 255 KB - mhealth_v8i5e15762_app2.png ]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1696 KB - mhealth_v8i5e15762_app3.pdf ]

## References


1. McEvoy MD, Field LC, Moore HE, Smalley JC, Nietert PJ, Scarbrough SH. The effect of adherence to ACLS protocols on survival of event in the setting of in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):82-87 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.019] [Medline: 24103233]
2. Link MS, Berkow LC, Kudenchuk PJ, Halperin HR, Hess EP, Moitra VK, et al. Part 7: Adult Advanced Cardiovascular Life Support: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015 Nov 3;132(18 Suppl 2):S444-S464. [doi: 10.1161/CIR.0000000000000261] [Medline: 26472995]
3. Khera R, Chan PS, Donnino M, Girotra S, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Hospital variation in time to epinephrine for nonshockable in-hospital cardiac arrest. Circulation 2016 Dec 20;134(25):2105-2114 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.116.025459] [Medline: 27908910]
4. Chan PS, Nichol G, Krumholz HM, Spertus JA, Nallamothu BK, American Heart Association National Registry of Cardiopulmonary Resuscitation (NRCPR) Investigators. Hospital variation in time to defibrillation after in-hospital cardiac arrest. Arch Intern Med 2009 Jul 27;169(14):1265-1273. [doi: 10.1001/archinternmed.2009.196] [Medline: 19636027]
5. Merchant RM, Berg RA, Yang L, Becker LB, Groeneveld PW, Chan PS, American Heart Association's Get With the Guidelines-Resuscitation Investigators. Hospital variation in survival after in-hospital cardiac arrest. J Am Heart Assoc 2014 Jan 31;3(1):e000400 [FREE Full text] [doi: 10.1161/JAHA.113.000400] [Medline: 24487717]
6. Hayes CW, Rhee A, Detsky ME, Leblanc VR, Wax RS. Residents feel unprepared and unsupervised as leaders of cardiac arrest teams in teaching hospitals: a survey of internal medicine residents. Crit Care Med 2007 Jul;35(7):1668-1672. [doi: 10.1097/01.CCM.0000268059.42429.39] [Medline: 17507825]
7. Lauridsen KG, Schmidt AS, Adelborg K, Løfgren B. Organisation of in-hospital cardiac arrest teams - a nationwide study. Resuscitation 2015 Apr;89:123-128. [doi: 10.1016/j.resuscitation.2015.01.014] [Medline: 25617486]
8. Field LC, McEvoy MD, Smalley JC, Clark CA, McEvoy MB, Rieke H, et al. Use of an electronic decision support tool improves management of simulated in-hospital cardiac arrest. Resuscitation 2014 Jan;85(1):138-142 [FREE Full text] [doi: 10.1016/j.resuscitation.2013.09.013] [Medline: 24056391]
9. Chan PS, Krumholz HM, Nichol G, Nallamothu BK, American Heart Association National Registry of Cardiopulmonary Resuscitation Investigators. Delayed time to defibrillation after in-hospital cardiac arrest. N Engl J Med 2008 Jan 3;358(1):9-17. [doi: 10.1056/NEJMoa0706467] [Medline: 18172170]
10. Donnino MW, Salciccioli JD, Howell MD, Cocchi MN, Giberson B, Berg K, American Heart Association's Get With The Guidelines-Resuscitation Investigators. Time to administration of epinephrine and outcome after in-hospital cardiac arrest with non-shockable rhythms: retrospective analysis of large in-hospital data registry. Br Med J 2014 May 20;348:g3028 [FREE Full text] [doi: 10.1136/bmj.g3028] [Medline: 24846323]
11. Patel KK, Spertus JA, Khariton Y, Tang Y, Curtis LH, Chan PS, American Heart Association's Get With the Guidelines–Resuscitation Investigators. Association between prompt defibrillation and epinephrine treatment with long-term survival after in-hospital cardiac arrest. Circulation 2018 May 8;137(19):2041-2051 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.030488] [Medline: 29279412]
12. Meaney PA, Bobrow BJ, Mancini ME, Christenson J, de Caen AR, Bhanji F, CPR Quality Summit Investigators, the American Heart Association Emergency Cardiovascular Care Committee, the Council on Cardiopulmonary, Critical Care, Perioperative and Resuscitation. Cardiopulmonary resuscitation quality: [corrected] improving cardiac resuscitation outcomes both inside and outside the hospital: a consensus statement from the American Heart Association. Circulation 2013 Jul 23;128(4):417-435. [doi: 10.1161/CIR.0b013e31829d8654] [Medline: 23801105]
13. Neumar RW, Shuster M, Callaway CW, Gent LM, Atkins DL, Bhanji F, et al. Part 1: Executive Summary: 2015 American Heart Association Guidelines Update for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation 2015 Nov 3;132(18 Suppl 2):S315-S367. [doi: 10.1161/CIR.0000000000000252] [Medline: 26472989]
14. Nelson KL, Shilkofski NA, Haggerty JA, Saliski M, Hunt EA. The use of cognitive AIDS during simulated pediatric cardiopulmonary arrests. Simul Healthc 2008;3(3):138-145. [doi: 10.1097/SIH.0b013e31816b1b60] [Medline: 19088657]


## Abbreviations

**ACLS:** Advanced Cardiovascular Life Support
**AHA:** American Heart Association
**CPR:** cardiopulmonary resuscitation
**IHCA:** in-hospital cardiac arrest
**PEA:** pulseless electrical activity

*Edited by G Eysenbach; submitted 03.08.19; peer-reviewed by M McEvoy, M Görges; comments to author 28.09.19; revised version received 21.11.19; accepted 04.02.20; published 19.05.20.*

*Please cite as:*
*Hejjaji V, Malik AO, Peri-Okonny PA, Thomas M, Tang Y, Wooldridge D, Spertus JA, Chan PS*
*Mobile App to Improve House Officers' Adherence to Advanced Cardiac Life Support Guidelines: Quality Improvement Study*
*JMIR Mhealth Uhealth 2020;8(5):e15762*
*URL: https://mhealth.jmir.org/2020/5/e15762*
*doi:10.2196/15762*
*PMID:32427115*

Original Paper

# The Personal Health Network Mobile App for Chemotherapy Care Coordination: Qualitative Evaluation of a Randomized Clinical Trial

Victoria Ngo[1], PhD; Cynthia G Matsumoto[1], MS; Jill G Joseph[1], MD, PhD; Janice F Bell[1], MN, MPH, PhD; Richard J Bold[2], MD; Andra Davis[3], RN, MN, PhD; Sarah C Reed[4], MSW, MPH, PhD; Katherine K Kim[1], MPH, MBA, PhD

[1]Betty Irene Moore School of Nursing, University of California, Davis, Sacramento, CA, United States

[2]Comprehensive Cancer Center, University of California Davis Health, Sacramento, CA, United States

[3]Washington State University College of Nursing – Vancouver, Vancouver, WA, United States

[4]Division of Social Work, California State University Sacramento, Sacramento, CA, United States

**Corresponding Author:**
Katherine K Kim, MPH, MBA, PhD
Betty Irene Moore School of Nursing
University of California, Davis
2450 48th Street, Suite 2600
Sacramento, CA
United States
Phone: 1 5107615461
Email: kathykim@ucdavis.edu

## Abstract

**Background:** Cancer care coordination addresses the fragmented and inefficient care of individuals with complex care needs. The complexity of care coordination can be aided by innovative technology. Few examples of information technology-enabled care coordination exist beyond the conventional telephone follow-up. For this study, we implemented a custom-designed app, the *Personal Health Network* (PHN)—a Health Insurance Portability and Accountability Act-compliant social network built around a patient to enable patient-centered health and health care activities in collaboration with clinicians, care team members, caregivers, and others designated by the patient. The app facilitates a care coordination intervention for patients undergoing chemotherapy.

**Objective:** This study aimed to understand patient experiences with PHN technology and assess their perspectives on the usability and usefulness of PHNs with care coordination during chemotherapy.

**Methods:** A two-arm randomized clinical trial was conducted to compare the PHN and care coordination with care coordination alone over a 6-month period beginning with the initiation of chemotherapy. A semistructured interview guide was constructed based on a theoretical framework of technology acceptance addressing usefulness, usability, and the context of use of the technology within the participant's life and health care setting. All participants in the intervention arm were interviewed on completion of the study. Interviews were recorded and transcribed verbatim. A summative thematic analysis was completed for the transcribed interviews. Features of the app were also evaluated.

**Results:** A total of 27 interviews were completed. The resulting themes included the care coordinator as a partner in care, learning while sick, comparison of other technology to make sense of the PHN, communication, learning, usability, and usefulness. Users expressed that the nurse care coordinators were beneficial to them because they helped them stay connected to the care team and answered their questions. They shared that the mobile app gave them access to the health information they were seeking. Users expressed that the mobile app would be more useful if it was fully integrated with the electronic health record.

**Conclusions:** The findings highlight the value of care coordination from the perspectives of cancer patients undergoing chemotherapy and the important role of technology, such as the PHN, in enhancing this process by facilitating better communication and access to information regarding their illness.

***(JMIR Mhealth Uhealth 2020;8(5):e16527)*** doi:10.2196/16527

**KEYWORDS**

care coordination, continuity of patient care; oncology; chemotherapy; patient-centered care; mobile health; technology adoption

# Introduction

## Background

Cancer care in the United States is fragmented [1-5] and complex [6]. The management of therapies, such as chemotherapy, requires the coordination of multiple hospital services and outpatient clinics [7]. Care coordination has been identified as a promising strategy for improving health care quality [8,9]. Engaging health care teams to actively participate in care coordination can be beneficial to patients in areas of improving communication, building trust, and facilitating transitions in care [10,11]. Successful care coordination involves effective communication among patients, their family members, and their care providers [12,13]. Communication between cancer patients and their health care team members can affect important health care decisions [14-16]. Patients face challenges such as lack of effective ways to document their health information while at the clinic or when they are away from home and lack of access care-related information [17].

It has been suggested that technology can aid in care coordination [18,19]. However, most information technology-enabled care coordination interventions have primarily utilized telephonic contact with limited examples of telehealth [9,20]. With technologies in communication and computing that have improved rapidly over the past decade, mobile health (mHealth) has enabled the collection of data, encouragement of healthy lifestyle changes, and improved management of care, especially in underserved and remote areas [21,22]. Mobile apps developed for cancer treatment can facilitate patient and provider communication, help manage patient information, and provide education around treatment follow-up [23]. Some features that may support a person's confidence in their ability to manage their own care include calendars, logging symptoms, tracking medications, and taking notes as needed [17]. Although mHealth is promising, the specific benefits to cancer care coordination have yet to be evaluated.

## Objective

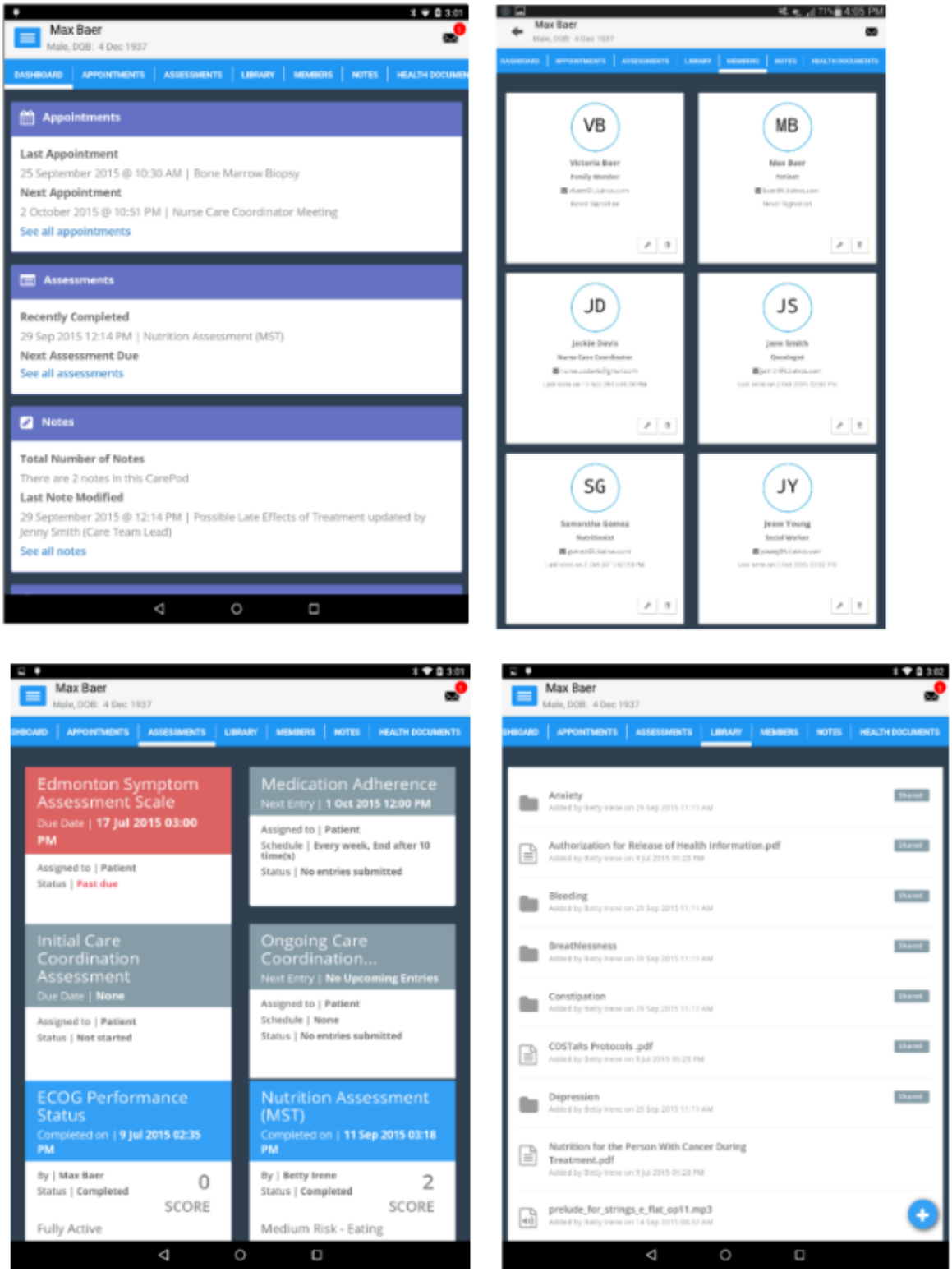
The Personal Health Network (PHN), a personalized, electronic social network built around a patient for collaboration with clinicians, care team members, caregivers, and others designated by a patient, was designed to address the challenges of cancer care coordination [24]. The objective of this study was to understand the participants' experiences with PHN technology and to assess their perceptions of usability and usefulness of the PHN on care coordination during chemotherapy.

# Methods

## Application Description

The PHN care coordination mHealth app was developed by a multidisciplinary group of experts who reviewed features from the published literature, assessed prototype versions [24] and conducted a user-centered design study with an evaluation of usability among patients [25]. The app consists of a dashboard for viewing all components of the care plan, contact information for members of the care team, regular symptom assessment surveys and other validated instruments to collect data on health issues and patient-reported outcomes, a self-management library of curated health information in Web, print, and video formats, a calendar with space for open-ended notes/journaling, secure messaging, and multiparty video chat (Figure 1).

**Figure 1.** The Personal Health Network app, top left, dashboard, top right, care team including health professionals and family caregivers, bottom left, symptom assessment and person-reported outcomes, bottom right, self-management library.




## Recruitment and Enrollment

This study was a component of a small, two-arm (N=63), randomized, pragmatic trial, in which the intervention group received the PHN technology and nurse care coordination, whereas the control group received nurse care coordination alone. Three registered nurses with care coordination training and experience provided care coordination to both arms. All trial participants were English-speaking, over the age of 18 years, received care at an urban comprehensive cancer center, had a primary diagnosis of cancer (any site), were initiating chemotherapy and had an expected survival of at least six months. All participants were followed up for 6 months after enrollment. Those in the intervention arm received an 8.4-inch Samsung Galaxy tablet with Wi-Fi and 4G data plan (Galaxy Tab Pro 8.4 SM-T325 loaded with Android 4.4 Kitkat and

TouchWiz UX software), and an individual orientation to the tablet and the PHN on enrollment. Technical assistance was embedded in the PHN app, and a telephone helpline was also made available. On study completion, all participants were allowed to retain the tablet. All participants in the intervention arm were asked to participate in the interviews. This study was approved by the Institutional Review Board of the University of California Davis. Written informed consent was obtained before data collection (trial registration: NCT02238951).

### Data Collection

Interviewers used a semistructured interview guide based on concepts of the Unified Theory of Acceptance and Use of Technology [26], including usability and usefulness, impact on health, the experience of the participants using the PHN, usability of features, and barriers and facilitators to use. Interviews were conducted with the intervention-arm participants at 6 months or at the end of care coordination, whichever was earlier, recorded, and transcribed verbatim. Interviews were conducted by three individuals trained by a senior investigator who also reviewed recordings and discussed interview techniques to reinforce shared understanding and consistency among them.

### Data Analysis

A two-phased thematic analysis method [27] was utilized: first, to identify the general themes and second, to review the themes identified in the context of usability and usefulness of the PHN app. NVivo 12 Pro (QSR International) was used to organize the data during the analysis. Two analysts (VN and CG) first familiarized themselves with the data by reading through the interview transcripts. They selected words and short phrases that symbolically evoked a salient attribute (ie, single idea codes) [27,28] and noted if the terms suggested positive, neutral, or negative sentiments. The researchers collaborated on the first two transcripts to develop a draft codebook, which was used for independent coding of the remaining transcripts. Any additions or revisions to the codes were discussed and added as needed. The analysts compared their coding and worked to align the differences found. Notations were made with analytic memos for discrepancies and changes in coding. Discrepancies between coders were resolved by a third researcher (KK). The three researchers conducted a final review of all the interviews to iteratively compare and discuss the patterns that were refined into themes based on discussion and consensus.

## Results

### Participant Characteristics

A total of 33 participants were randomized to the intervention arm of the underlying trial, and 82% (27/33) participated in interviews (Table 1). One participant passed away from the disease while enrolled in the study, and the rest declined being interviewed because of scheduling conflicts. The mean age was 59 years (range 22-79 years). Most participants were female (23/27, 85%) and white (24/27, 89%). The participants were highly educated (17/27, 63% college graduates), and more than a third had high incomes (10/27, 37%, had annual incomes >US $80,000).

**Table 1.** Characteristics of interview participants.

| Variables | Values, n (%) |
|---|---|
| **Gender** | |
| Male | 4 (15) |
| Female | 23 (85) |
| **Age (years)** | |
| 18 to 45 | 1 (4) |
| 46 to 64 | 15 (56) |
| 65 and older | 11 (41) |
| **Race and ethnicity** | |
| Hispanic or Latino | 0 (0) |
| Non-Hispanic white | 24 (89) |
| Black or African American | 0 (0) |
| Asian | 2 (7) |
| Native Hawaiian or Pacific Islander | 1 (4) |
| **Education** | |
| High school graduate or GED[a] | 2 (7) |
| Some college | 8 (30) |
| College graduate | 6 (22) |
| More than 4-year college degree | 11 (41) |
| **Income** | |
| Less than US $49,999 | 7 (26) |
| US $50,000 to US $79,999 | 9 (33) |
| US $80,000 and more | 10 (37) |
| Prefer not to state | 1 (4) |

[a]GED: general education diploma.

## Thematic Analysis of the Interviews

A total of 82 single idea codes were generated and separated into positive and negative subgroups, resulting in 177 unique codes (13 codes were neutral). Key themes were uncovered through the systematic categorization of codes. In addition to overall usability and overall usefulness, 5 themes emerged from the data.

The 5 themes are listed below, and exemplar quotes are provided in Multimedia Appendix 1.

- Nurse care coordinator as a partner in care: this theme referred to someone who had a relationship with the patient, routinely checked in with the patient, helped find resources, communicated between team members, and assisted with problem solving;
- Learning: the learning theme refers to learning how to use a mobile app both via teaching and via experience using it;
- Learning while sick: participants gave insight on what it felt like to learn something new while going through chemotherapy, and for some, experiencing side effects from treatment;
- Comparison of other technology to make sense of the PHN: many participants made references to similar technologies, both apps, and devices, they used and how this knowledge was transferrable and helped them make sense of the PHN;
- Communication: the communication theme encompasses both access to people and information.

### *Overall Usability*

Participants expressed that a usable app would be accessible through the internet, compatible with other apps (such as the electronic health record patient portal), easy to use, portable, navigable, and performed quickly. Some participants appreciated using the PHN because of access to information:

> *You know I never tried to manage my healthcare through PHN or even if I can pull MyChart into it or any of those things before. So it's all been a new and very good experience to always have access. The accessibility is just the best part of the whole thing. It's a 24/7 type thing.*

The portability of the tablet and its ease of use allowed the participants to use the technology:

> *And I know that when I came here for treatment and used it, I liked using it. It was you know very small, not heavy, so it's easy to carry with you. And like I said it's relatively easy to use too.*

Connectivity to the internet was important:

> *I think everybody you know I think if you're going to push through with this program I think it'll be very helpful for all patients as long as they have access to Internet.*

Not everyone found the app easy to use or usable because of the slow performance they experienced:

> *And I have very, very fast internet connections and things were sluggish and kind of kludgy. I know that's not a technological term but I was challenged and I'm looking at this thing going, this is literally driving me crazy. All this, you know, open up my laptop and go find this information quickly somewhere else.*

Another participant said:

> *I know it's asking a lot but that was just my initial feel because I find the other technology so easy to use and I did not find that with the one that we were using from my care for the Project.*

Due to some issues with usability and software upgrades, some participants felt discouraged from using the PHN:

> *And I'm trying to think back over the last 6-8 months since I've been using it and it just seems like it's just been really glitchy. And there has been a lot of changes so you know it seems like every time I was starting it up, I was looking at another update or another change.*
>
> *And that's sort of what it felt like for me, it's like I mean I literally had to put on my glasses and to even see the font sizes and the buttons. And those are just navigational issues that I think that start that frustration where you go, I don't care what else is in here because I can't even get past the screen opening, or I can't do those kinds of things. And I think the organization of it was confusing and it didn't seem to make a whole lot of sense to me.*

### *Overall Usefulness*

Usefulness refers to the benefits of technology for participants to accomplish their health goals. Participants generally expressed that the PHN was useful because of the ability to answer questions beyond regular business hours. One person said:

> *It was helpful because it was on my time. So when I had to come up with a question at 7 o'clock at night when there's no one there and come 7 o'clock in the morning when I'm really running around doing something else and maybe not having the time to sit down and think about it at that moment I was able to ask my question whenever I wanted knowing I wasn't going to get a response till the next day but at least I wasn't, you know I was able to deal with it then before I forgot, before something else happened. So it's really–it's being able to do things on my time and my schedule.*

Another participant mentioned that the PHN was helpful at the beginning of the chemotherapy journey, but not so much later:

> *Well, when I was really sick when I was in the beginning, I used to a lot more than I of course use now because I don't you know I'm not using it for the things that the nurse coordinator was dropping in for me and for that sort of stuff it was really, really useful.*
>
> *As time progressed over a six month period I didn't use it as much. I didn't find it as useful because I wasn't searching out for those answers.*

Another participant wanted more interaction with the PHN:

> *I was hoping it to be more interactive and more personalized to me as opposed to it was kind of generic, the information that was sent.*

A suggestion mentioned was that the PHN might be useful for patients who require more support:

> *If I was in a situation where I needed more support it would have been really well, and so I think finding out how much support does the patient want and expect. Because I had the support at home, I had other things there I didn't feel that this was something that I needed but I can see where there are people that this would fill a gap in their care and I think it would be very well. And I'm probably not the best person to fully utilize the benefits that you have there.*

## Usefulness of the Personal Health Network's Functions

Overall, participants were more positive than negative regarding PHN functions. Table 2 shows the number of positive and negative comments made for each function. Participants identified ways in which these functions were useful to their overall health goals or specific needs related to chemotherapy care. Examples of usefulness are summarized in Multimedia Appendix 2.

**Table 2.** Perception of the Personal Health Network by function.

| Function | Positive comments, n | Negative comments, n | Neutral comments, n | Total, n |
|---|---|---|---|---|
| Library | 72 | 12 | 4 | 88 |
| Survey | 30 | 18 | 0 | 48 |
| Messaging | 34 | 11 | 0 | 45 |
| Camera | 1 | 6 | 0 | 7 |
| Calendar | 4 | 8 | 19 | 31 |
| Overall interface | 1 | 8 | 0 | 9 |

# Discussion

## Comparison With Prior Work

As digital technology continues to develop and create more opportunities to provide health worldwide [29], the implementation of a digital health ecosystem—where the community health network of people, devices, and technology are interconnected—must take into consideration not only the interactions of technologies but also the network and interaction of key health care stakeholders both in receiving and providing care [30,31]. Especially in digital health ecosystems within the care and assistance domain, not all stakeholders are often invited to participate in the technology design process [32]. There have been few randomized controlled trials (RCTs) of mHealth in cancer care coordination described in the literature, and fewer still that involved clinicians and patients in the design and testing of the technology [33]. Before this study, the research team proposed a conceptual framework for person-centered, community-wide care coordination and defined the concept of *point of need* for coordination, which includes both settings where health care services are delivered and everyday settings where individuals need to make health-related decisions [9,34]. In addition, a user-centered design study of the PHN prototype investigated the usability of the platform to improve the design before starting the RCT [25].

## Principal Results

The work presented in this paper investigates both usability and usefulness upon completion of care coordination among individuals undergoing chemotherapy. Similar to other studies [33,35,36], our findings show that adults find mobile apps useful for monitoring symptoms and side effects, and as a way of communicating needs and coordinating care in a timely manner [37]. Although the literature on cancer care coordination activities in the United States is sparse [38], the themes that emerged from our summative interviews contribute viewpoints that can enhance future interventions and the design and implementation of mobile apps for this purpose.

An important aspect of care coordination is access to health information. Individuals with cancer and their caregivers want information about the illness, treatment options, care needs, and often turn to the internet to seek resources [15,39-41]. Individuals with cancer face challenges in managing their health-related information [17], which includes collecting relevant data, communicating about that data, and accessing informational resources to make sense of the data. There can be barriers to access, such as paywalls or membership-only portals [39]. Moreover, individuals may experience information overload, where synthesizing information becomes an obstacle [42,43].

mHealth is well suited to facilitate information management tasks and may support health impacts [44]. The PHN, which was designed with these information management tasks in mind, was viewed as a supportive platform for accomplishing them. Although having the PHN was helpful for organizing information and care, the presence of a nurse care coordinator was important to help with problem solving. The PHN complemented the knowledge and experience of a nurse care coordinator to help guide, organize, and tailor information. The PHN library feature, in particular, received a substantial proportion of positive sentiments, indicating that it was viewed as very beneficial to participants. Participants shared that it was most helpful when the nurse care coordinator worked with the patient to identify specific information in the library and highlight it on the participant's dashboard. It may be that the PHN library offered benefits in reducing this overload by curating relevant and clinician-endorsed information.

PHN symptom surveys were designed to increase awareness of symptoms, communicate about symptoms for early intervention, and track progress based on the patient-reported outcomes. Participants found that filling out the PHN symptom surveys was a simple task. They indicated that these surveys prompted them to think about their symptoms and discuss them with the nurse care coordinator who could deliver useful self-management information via the PHN or alert a physician for possible changes in therapy. This finding is aligned with previous studies that reported that tracking and reporting symptoms cause patients to reflect on their own well-being [45]. For those who were undergoing chemotherapy, tracking symptoms in real time increased awareness of self-care and improved communication with the health care team [46]. This also paves the way for the potential to improve clinical outcomes. Basch et al [47] reported that an intervention using Web-based symptom collection, evaluated in a large RCT among patients with cancer, was associated with improved outcomes, including quality-adjusted survival rates, fewer emergency department visits and hospitalizations, and improved quality of life. Thus, the combination of symptom awareness, self-management support, and early intervention shows promise in improving both clinical and person-centered outcomes in cancer care.

PHN messaging features were also viewed positively. Participants expressed the importance of communication outside of regular business operating hours. One participant highlighted the stress caused by not being able to communicate with a member of the health care team when experiencing an unfamiliar symptom. Although messages in the PHN were not monitored outside of business hours, there may have been some comfort in being able to express concern at the moment with the confidence that the care coordinator would respond the next morning. In our on-demand information era, consumers have become used to a very quick turnaround on questions and concerns, and this expectation has added urgency when involving a health care concern.

Although participants were willing to use the PHN, difficulties associated with learning something new while sick was a reality. Participants experiencing *chemo brain*—a term participants used to describe how they feel their thinking is impaired during chemotherapy treatment—emerged as a challenge to adopting new technology or intervention. Even though the need for information resources during initial diagnosis and early treatment may be great, the ability to adopt a tool such as the PHN may be difficult. Future study designs might target caregivers who are actively involved in care coordination. The PHN may also be an aid to survivors who are managing maintenance therapy or a survivorship care plan. In addition, participants did not use the tablet's built-in camera, but some were enthusiastic about trying to show their doctor *something* (they found concerning on their body) or video chat face-to-face with a care team member. A future study may consider the different preferences of communication routes (eg, telephone, messaging, video conferencing, or virtual reality) when coordinating care.

Perceptions of the usability of the PHN were mixed. Participants made suggestions for improvements in navigation within the app, visual layout and increased font size and graphics, confirmation of tasks completed, and reminders for upcoming tasks. Even in the current environment of ubiquitous access to broadband, participants still reported challenges connecting to the internet (a data plan was provided). Interoperability with the patient portal in the electronic health record was highly preferred. There is room to improve technology to further enhance adoption.

### Limitations

There were several limitations to this study. Participants were recruited from one urban cancer center and were primarily older, white females with a higher socioeconomic status. Interviews were conducted as each participant completed their study period, and the sample size was determined based on an RCT. Coding was conducted after all the interviews were complete. We were not able to add interviews to assure theoretical saturation or explore new avenues of inquiry. Thus, our analysis offers limited perspectives on the usefulness and usability of PHNs.

### Conclusions

This study contributed to expanding the knowledge of cancer care coordination efforts, specifically around incorporating the use of technology to organize information, services, and people. Insight into the patient experience of PHN during chemotherapy provided a better understanding of participants' perceptions of usability and usefulness. Findings from this analysis revealed that participants believed that care coordination is a valuable benefit for cancer patients undergoing chemotherapy, and the use of PHN technology can enhance this process by facilitating better communication and access to information.

## Acknowledgments

This study was funded by the McKesson Foundation (JJ and KK, principal investigators), a grant from Boston University Center for the Future of Technology in Cancer Care (KK, principal investigator) via the National Institute of Biomedical Imaging and Bioengineering Award U54EB015403 (Catherine M Klapperich, center principal investigator), and a Gordon and Betty Moore Foundation grant to the Betty Irene Moore School of Nursing to fund graduate student researchers (VN and CM). The authors thank Chelsie Marie Gonzales Antonio, RN, Thuy B Le, and Ronald Grummer for help with the recruitment of participants for this study. The authors thank Phillip Bautista and Nicole Mahr for assisting with the participant interviews. Finally, the authors thank Drs Theresa H Keegan, Brian A Jonas, Elizabeth I Rice, and Sheryl L Catz for providing thoughtful feedback on drafts of the manuscript.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Exemplar quotations for themes.
[DOCX File , 29 KB - mhealth_v8i5e16527_app1.docx ]

Multimedia Appendix 2
Exemplar quotations of the usefulness of the Personal Health Network features.
[DOCX File , 28 KB - mhealth_v8i5e16527_app2.docx ]

## References

1. Institute of Medicine, Board on Health Care Services, Committee on Improving the Quality of Cancer Care: Addressing the Challenges of an Aging Population. In: Levit L, Balogh E, Nass S, Ganz PA, editors. Delivering High-quality Cancer Care: Charting A New Course For A System In Crisis. Washington, DC: National Academies Press; 2014.
2. Ganz PA, Levit LA. Charting a new course for the delivery of high-quality cancer care. J Clin Oncol 2013 Dec 20;31(36):4485-4487. [doi: 10.1200/jco.2013.53.7993]
3. Ganz PA. Delivering patient-centered care in the setting of advanced cancer: what does a clinical risk-prediction model have to do with it? JAMA Oncol 2015 Jul 1;1(4):430-432. [doi: 10.1001/jamaoncol.2015.0832] [Medline: 26181248]
4. Balogh EP, Ganz PA, Murphy SB, Nass SJ, Ferrell BR, Stovall E. Patient-centered cancer treatment planning: improving the quality of oncology care. Summary of an institute of medicine workshop. Oncologist 2011 Nov 29;16(12):1800-1805 [FREE Full text] [doi: 10.1634/theoncologist.2011-0252] [Medline: 22128118]
5. Schoen C, Osborn R, Squires D, Doty M, Pierson R, Applebaum S. New 2011 survey of patients with complex care needs in eleven countries finds that care is often poorly coordinated. Health Aff (Millwood) 2011 Dec;30(12):2437-2448. [doi: 10.1377/hlthaff.2011.0923] [Medline: 22072063]
6. Institute of Medicine, Committee on Quality of Health Care in America. Crossing The Quality Chasm: A New Health System For The 21st Century. Washington, DC: National Academies Press; 2001.
7. Lamé G, Jouini O, Stal-Le Cardinal J. Outpatient chemotherapy planning: a literature review with insights from a case study. IIE Trans Healthc Syst Eng 2016 May 31;6(3):127-139. [doi: 10.1080/19488300.2016.1189469]
8. US Department of Health and Human Services. Agency for Healthcare Research and Quality. 2013. 2013 Annual Progress Report to Congress: National Strategy for Quality Improvement in Health Care URL: https://www.ahrq.gov/workingforquality/reports/2013-annual-report.html [accessed 2020-01-01]
9. Kim KK, Bell JF, Reed SC, Whitney RL. Coordination at the point of need. In: Hesse BW, Ahern D, Beckjord E, editors. Oncology Informatics: Using Health Information Technology To Improve Processes And Outcomes In Cancer. Cambridge, Massachusetts: Academic Press; 2016:81-100.
10. Vandyk AD, Harrison MB, Macartney G, Ross-White A, Stacey D. Emergency department visits for symptoms experienced by oncology patients: a systematic review. Support Care Cancer 2012 Aug;20(8):1589-1599. [doi: 10.1007/s00520-012-1459-y] [Medline: 22526151]
11. Peikes D, Chen A, Schore J, Brown R. Effects of care coordination on hospitalization, quality of care, and health care expenditures among medicare beneficiaries: 15 randomized trials. J Am Med Assoc 2009 Feb 11;301(6):603-618. [doi: 10.1001/jama.2009.126] [Medline: 19211468]
12. Arora NK, Jensen RE, Sulayman N, Hamilton AS, Potosky AL. Patient-physician communication about health-related quality-of-life problems: are non-Hodgkin lymphoma survivors willing to talk? J Clin Oncol 2013 Nov 1;31(31):3964-3970 [FREE Full text] [doi: 10.1200/JCO.2012.47.6705] [Medline: 24062408]
13. Walsh J, Young J, Harrison J, Butow P, Solomon M, Masya L, et al. What is important in cancer care coordination? A qualitative investigation. Eur J Cancer Care (Engl) 2011 Mar;20(2):220-227. [doi: 10.1111/j.1365-2354.2010.01187.x] [Medline: 20477854]
14. Hoerger M, Epstein RM, Winters PC, Fiscella K, Duberstein PR, Gramling R, et al. Values and options in cancer care (VOICE): study design and rationale for a patient-centered communication and decision-making intervention for physicians, patients with advanced cancer, and their caregivers. BMC Cancer 2013 Apr 9;13(1):188 [FREE Full text] [doi: 10.1186/1471-2407-13-188] [Medline: 23570278]
15. Hesse BW, Beckjord E, Rutten LJ, Fagerlin A, Cameron LD. Cancer communication and informatics research across the cancer continuum. Am Psychol 2015;70(2):198-210. [doi: 10.1037/a0036852] [Medline: 25730725]
16. Beckjord EB, Arora NK, McLaughlin W, Oakley-Girvan I, Hamilton AS, Hesse BW. Health-related information needs in a large and diverse sample of adult cancer survivors: implications for cancer care. J Cancer Surviv 2008 Sep;2(3):179-189. [doi: 10.1007/s11764-008-0055-0] [Medline: 18792791]
17. Klasnja P, Hartzler A, Powell C, Pratt W. Supporting cancer patients' unanchored health information management with mobile technology. AMIA Annu Symp Proc 2011;2011:732-741 [FREE Full text] [Medline: 22195130]
18. Hesse BW, Hanna C, Massett HA, Hesse NK. Outside the box: will information technology be a viable intervention to improve the quality of cancer care? J Natl Cancer Inst Monogr 2010;2010(40):81-89 [FREE Full text] [doi: 10.1093/jncimonographs/lgq004] [Medline: 20386056]
19. Mortimer L, Strawbridge L, Lukens E, Bassano A, Conway P, Kline R. Centers for Medicare & Medicaid Services' oncology care model: delivering higher value cancer care. Clin Pharmacol Ther 2017 May;101(5):569-571. [doi: 10.1002/cpt.638] [Medline: 28139833]
20. Dickinson R, Hall S, Sinclair JE, Bond C, Murchie P. Using technology to deliver cancer follow-up: a systematic review. BMC Cancer 2014 May 3;14:311 [FREE Full text] [doi: 10.1186/1471-2407-14-311] [Medline: 24885758]
21. Akter S, Ray P. mHealth - an ultimate platform to serve the unserved. Yearb Med Inform 2010;19(1):94-100. [doi: 10.1055/s-0038-1638697] [Medline: 20938579]
22. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, et al. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS Med 2013;10(1):e1001362 [FREE Full text] [doi: 10.1371/journal.pmed.1001362] [Medline: 23349621]

23. Davis SW, Oakley-Girvan I. mHealth education applications along the cancer continuum. J Cancer Educ 2015 Jun;30(2):388-394. [doi: 10.1007/s13187-014-0761-4] [Medline: 25482319]
24. Kim K, Bell J, Reed S, Joseph J, Bold R, Cerrone K, et al. A Novel Personal Health Network for Patient-centered Chemotherapy Care Coordination. In: Proceedings of the International Conference on Collaboration Technologies and Systems. 2014 Presented at: CTS'14; May 19-23, 2014; Minneapolis, MN, USA. [doi: 10.1109/cts.2014.6867603]
25. Kim K, Bell J, Bold R, Davis A, Ngo V, Reed S, et al. A personal health network for chemotherapy care coordination: evaluation of usability among patients. Stud Health Technol Inform 2016;225:232-236. [doi: 10.3233/978-1-61499-658-3-232] [Medline: 27332197]
26. Venkatesh V, Thong JY, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. Manag Inf Syst Q 2012;36(1):157. [doi: 10.2307/41410412]
27. Saldaña J. The Coding Manual For Qualitative Researchers. London: Sage Publications Ltd; 2015.
28. Miles M, Huberman A, Saldaña J. Qualitative Data Analysis: A Methods Sourcebook. Thousand Oaks, California: Sage Publications, Inc; 2014.
29. León MC, Nieto-Hipólito JI, Garibaldi-Beltrán J, Amaya-Parra G, Luque-Morales P, Magaña-Espinoza P, et al. Designing a model of a digital ecosystem for healthcare and wellness using the business model canvas. J Med Syst 2016 Jun;40(6):144. [doi: 10.1007/s10916-016-0488-3] [Medline: 27118010]
30. Iyawa GE, Herselman M, Botha A. Digital health innovation ecosystems: from systematic literature review to conceptual framework. Procedia Comput Sci 2016;100:244-252. [doi: 10.1016/j.procs.2016.09.149]
31. Iyawa GE, Herselman M, Botha A. A Scoping Review of Digital Health Innovation Ecosystems in Developed and Developing Countries. In: Proceedings of the 2017 IST-Africa Week Conference. 2017 Presented at: IST-Africa'17; May 30-June 2, 2017; Windhoek, Namibia. [doi: 10.23919/istafrica.2017.8102325]
32. Marcos-Pablos S, García-Peñalvo FJ. Technological ecosystems in care and assistance: a systematic literature review. Sensors (Basel) 2019 Feb 9;19(3):708 [FREE Full text] [doi: 10.3390/s19030708] [Medline: 30744096]
33. Davis SW, Oakley-Girvan I. Achieving value in mobile health applications for cancer survivors. J Cancer Surviv 2017 Aug;11(4):498-504. [doi: 10.1007/s11764-017-0608-1] [Medline: 28342093]
34. Kim KK, Bell JF, Boicey C, Freeman-Daily J, Hull S, McCollister-Slipp A, et al. A Framework for Person-Centered, Community-Wide Care Coordination. In: Proceedings of the 2015 AMIA Annual Symposium Proceedings Archive. Washington, DC: American Medical Informatics Association; 2015 Presented at: AMIA'15; November 14-18, 2015; San Francisco, California p. 78-79 URL: https://knowledge.amia.org/59310-amia-1.2741865/t001-1.2745946/f001-1.2745947/2247697-1.2746098/2247697-1.2746099?qr=1
35. Maguire R, Ream E, Richardson A, Connaghan J, Johnston B, Kotronoulas G, et al. Development of a novel remote patient monitoring system: the advanced symptom management system for radiotherapy to improve the symptom experience of patients with lung cancer receiving radiotherapy. Cancer Nurs 2015;38(2):E37-E47. [doi: 10.1097/NCC.0000000000000150] [Medline: 24836956]
36. Larsen M, Rowntree J, Young A, Pearson S, Smith J, Gibson O, et al. Chemotherapy side-effect management using mobile phones. Conf Proc IEEE Eng Med Biol Soc 2008;2008:5152-5155. [doi: 10.1109/IEMBS.2008.4650374] [Medline: 19163877]
37. Baseman J, Revere D, Baldwin L. A mobile breast cancer survivorship care app: pilot study. JMIR Cancer 2017 Sep 26;3(2):e14 [FREE Full text] [doi: 10.2196/cancer.8192] [Medline: 28951383]
38. Gorin SS, Haggstrom D, Han PKJ, Fairfield KM, Krebs P, Clauser SB. Cancer care coordination: a systematic review and meta-analysis of over 30 years of empirical studies. Ann Behav Med 2017 Aug;51(4):532-546. [doi: 10.1007/s12160-017-9876-2] [Medline: 28685390]
39. Papadakos J, Trang A, Cyr AB, Abdelmutti N, Giuliani ME, Snow M, et al. Deconstructing cancer patient information seeking in a consumer health library toward developing a virtual information consult for cancer patients and their caregivers: a qualitative, instrumental case study. JMIR Cancer 2017 May 24;3(1):e6 [FREE Full text] [doi: 10.2196/cancer.6933] [Medline: 28539305]
40. Hesse BW, Arora NK, Beckjord EB, Rutten LJ. Information support for cancer survivors. Cancer 2008 Jun 1;112(11 Suppl):2529-2540 [FREE Full text] [doi: 10.1002/cncr.23445] [Medline: 18428201]
41. Roach A, Lykins E, Gochett C, Brechting E, Graue L, Andrykowski M. Differences in cancer information-seeking behavior, preferences, and awareness between cancer survivors and healthy controls: a national, population-based survey. J Cancer Educ 2009;24(1):73-79 [FREE Full text] [doi: 10.1080/08858190802664784] [Medline: 19259869]
42. Mukherjee A, Bawden D. Health information seeking in the information society. Health Info Libr J 2012 Sep;29(3):242-246 [FREE Full text] [doi: 10.1111/j.1471-1842.2012.00997.x] [Medline: 22925387]
43. Clarke MA, Moore JL, Steege LM, Koopman RJ, Belden JL, Canfield SM, et al. Health information needs, sources, and barriers of primary care patients to achieve patient-centered care: a literature review. Health Informatics J 2016 Dec;22(4):992-1016. [doi: 10.1177/1460458215602939] [Medline: 26377952]
44. Fiordelli M, Diviani N, Schulz PJ. Mapping mhealth research: a decade of evolution. J Med Internet Res 2013 May 21;15(5):e95 [FREE Full text] [doi: 10.2196/jmir.2430] [Medline: 23697600]

45. Langius-Eklöf A, Christiansen M, Lindström V, Blomberg K, Nyman MH, Wengström Y, et al. Adherence to report and patient perception of an interactive app for managing symptoms during radiotherapy for prostate cancer: descriptive study of logged and interview data. JMIR Cancer 2017 Oct 31;3(2):e18 [FREE Full text] [doi: 10.2196/cancer.7599] [Medline: 29089290]
46. Breen S, Ritchie D, Schofield P, Hsueh Y, Gough K, Santamaria N, et al. The Patient Remote Intervention and Symptom Management System (PRISMS) - a Telehealth- mediated intervention enabling real-time monitoring of chemotherapy side-effects in patients with haematological malignancies: study protocol for a randomised controlled trial. Trials 2015 Oct 19;16:472 [FREE Full text] [doi: 10.1186/s13063-015-0970-0] [Medline: 26481873]
47. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]

## Abbreviations

**mHealth:** mobile health
**PHN:** Personal Health Network
**RCT:** randomized controlled trial

*Edited by G Eysenbach; submitted 07.12.19; peer-reviewed by GE Iyawa, T Muto; comments to author 19.01.20; revised version received 30.01.20; accepted 01.02.20; published 26.05.20.*

*Please cite as:*
*Ngo V, Matsumoto CG, Joseph JG, Bell JF, Bold RJ, Davis A, Reed SC, Kim KK*
*The Personal Health Network Mobile App for Chemotherapy Care Coordination: Qualitative Evaluation of a Randomized Clinical Trial*
*JMIR Mhealth Uhealth 2020;8(5):e16527*
*URL: http://mhealth.jmir.org/2020/5/e16527/*
*doi:10.2196/16527*
*PMID:32452814*

Original Paper

# The Most-Cited Authors Who Published Papers in JMIR mHealth and uHealth Using the Authorship-Weighted Scheme: Bibliometric Analysis

Wei-Chih Kan[1,2], MD; Willy Chou[3,4], MD; Tsair-Wei Chien[5], MBA; Yu-Tsen Yeh[6], BA; Po-Hsin Chou[7,8], PhD

[1]Department of Nephrology, Chi Mei Medical Center, Taiwan, Tainan, Taiwan
[2]Department of Biological Science and Technology, Chung Hwa University of Medical Technology, Tainan, Taiwan
[3]Department of Physical Medicine and Rehabilitation, Chi Mei Medical Center, Tainan, Taiwan
[4]Department of Physical Medicine and Rehabilitation, Chung Shan Medical University, Taichun, Taiwan
[5]Department of Medical Research, Chi Mei Medical Center, Taiwan, Tainan, Taiwan
[6]Medical School, St George's, University of London, London, United Kingdom
[7]Department of Orthopedics and Traumatology, Taipei Veterans General Hospital, Taipei, Taiwan
[8]School of Medicine, National Yang-Ming University, Taipei, Taiwan

**Corresponding Author:**
Po-Hsin Chou, PhD
School of Medicine
National Yang-Ming University
18F, 201, Section 2, Shipai Road, Beitou District
Taipei, 112
Taiwan
Phone: 886 228757557
Email: choupohsin@gmail.com

## Abstract

**Background:** Many previous papers have investigated most-cited articles or most productive authors in academics, but few have studied most-cited authors. Two challenges are faced in doing so, one of which is that some different authors will have the same name in the bibliometric data, and the second is that coauthors' contributions are different in the article byline. No study has dealt with the matter of duplicate names in bibliometric data. Although betweenness centrality (BC) is one of the most popular degrees of density in social network analysis (SNA), few have applied the BC algorithm to interpret a network's characteristics. A quantitative scheme must be used for calculating weighted author credits and then applying the metrics in comparison.

**Objective:** This study aimed to apply the BC algorithm to examine possible identical names in a network and report the most-cited authors for a journal related to international mobile health (mHealth) research.

**Methods:** We obtained 676 abstracts from Medline based on the keywords "JMIR mHealth and uHealth" (Journal) on June 30, 2018. The author names, countries/areas, and author-defined keywords were recorded. The BCs were then calculated for the following: (1) the most-cited authors displayed on Google Maps; (2) the geographical distribution of countries/areas for the first author; and (3) the keywords dispersed by BC and related to article topics in comparison on citation indices. Pajek software was used to yield the BC for each entity (or node). Bibliometric indices, including h-, g-, and x-indexes, the mean of core articles on g(Ag)=sum (citations on g-core/publications on g-core), and author impact factor (AIF), were applied.

**Results:** We found that the most-cited author was Sherif M Badawy (from the United States), who had published six articles on JMIR mHealth and uHealth with high bibliometric indices (h=3; AIF=8.47; x=4.68; Ag=5.26). We also found that the two countries with the highest BC were the United States and the United Kingdom and that the two keyword clusters of mHealth and telemedicine earned the highest indices in comparison to other counterparts. All visual representations were successfully displayed on Google Maps.

**Conclusions:** The most cited authors were selected using the authorship-weighted scheme (AWS), and the keywords of mHealth and telemedicine were more highly cited than other counterparts. The results on Google Maps are novel and unique as knowledge concept maps for understanding the feature of a journal. The research approaches used in this study (ie, BC and AWS) can be applied to other bibliometric analyses in the future.

*(JMIR Mhealth Uhealth 2020;8(5):e11567)* doi:10.2196/11567

**KEYWORDS**

betweenness centrality; authorship collaboration; Google Maps; social network analysis; knowledge concept map; the author-weighted scheme

## Introduction

### Background

As of April 12, 2018, more than 146 papers were found by the keyword "author collaboration" (Title), 1168 by "author collaboration," and 53 by "author collaboration" and "bibliometric" in the Medline Library. A phenomenal increase has been found in the number of research papers with multiple authors [1]. The knowledge of discovery is no longer contained merely in the departments of a local university but in an international article author byline [2]. Increasing academic pressure and prestige-concerned individuals with prolific publications have also been forced to claim authorship for many aspirants on paper publications [3]. Given academic developments in recent years, the features of author collaboration on one topic or for a specific journal should be investigated.

### Issue of Duplicate Authors in a Network

An author's publication features can be determined by social network analysis (SNA) [4-8]. However, no study currently in the literature describes the issue of duplicate names in bibliometric data, which might result in biases because some different authors with the same name exist [7]. For instance, authors [7] stressed that:

> *[T]here might be some biases of understanding for author collaboration because some different authors with the same name or abbreviation exist, who are affiliated to different institutions. The result of author relationship analysis for mHealth research would be influenced by the accuracy of the indexing author.*

Three main centrality measures (ie, degree, closeness, and betweenness) are frequently used to evaluate the influence (or power) momentum of an entity (or the author of a study) in a network [9,10]. Few studies have applied betweenness centrality (BC) to interpreting a network's characteristics. In this study, we aimed to explore whether BC can solve the problem of detecting duplicate authors in a network.

### Issue of Most-Cited Authors in a Given Journal

As of June 31, 2020, over 269 articles were found by searching the keyword "most cited" (Title) in PubMed Central (PMC) and 39 papers by "most productive author" or "most prolific author." However, few had studied most-cited authors. The reason might be that there is no quantitative scheme that has been successfully used to calculate weighted author credits in the literature; even many counting schemes have been proposed for quantifying coauthor contributions [11-13]. Thus, an authorship-weighted scheme (AWS) will be required for application to bibliometric metrics to allow for comparison.

### Issue of a Dashboard Possibly Shown on Google Maps

The author's publication patterns are always presented with static .jpg format pictures [4-7] instead of a dynamic dashboard that allows readers to see further details on their own. We have observed many bibliometric studies [7,14-19] using coword (or coauthor) analysis to visualize study data. However, no work has displayed their findings with a zoom-in and zoom-out functionality on Google Maps [20,21]. A breakthrough in showing data on Google Maps is a worthwhile task to develop.

### Objectives

The journal of JMIR mHealth and uHealth was targeted for BC algorithm application to examine possible duplicate authors with the same names in a network. Our goal is to select the most highly cited authors in author collaborations. Also, both features (ie, the affiliation regions distributed for the first author in geography, and the keywords related to article topics) will be investigated using the citation analysis in this study.

## Methods

### Data Collection

When searching the PubMed database (Pubmed.org) maintained by the US National Library of Medicine, we used the keywords "JMIR mHealth and uHealth" (Journal) on June 30, 2018. We then downloaded 676 articles that had been published since 2013, because the first article in JMIR mHealth and uHealth was published in 2013. An author-made Microsoft Excel (Microsoft Corporation, Albuquerque, New Mexico, United States) VBA (visual basic for applications) module was used to analyze the research data. All downloaded abstracts were based on the type of journal article involved. Ethical approval was not necessary for this study because all the data were obtained online from the Medline library.

### Social Network Analysis and the Betweenness Centrality

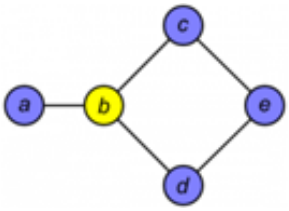
SNA [22] was applied to explore the pattern of entities in a system using the software Pajek [in Koeln; PajekMan in Osoje (Ossiach, Austria)] [23]. In keeping with the Pajek guidelines, we defined an author (or paper keyword) as a node (or an actor) that is connected to other nodes through the edge (or the relation). The number of connections usually defines the weight between two nodes.

Centrality is a vital index for analyzing a network. Any individual or keyword in the center of a social network will determine its influence on the network and its speed at gaining information [9,24]. In this study, we used the BC, which may be defined loosely as the number of times a node needs a given node to reach another node [9,25], as in, the number of shortest paths passing through a given node. The BC is expressed as follows, in Standalone Equation 1:

By contrast, the BC of node v, which is denoted as g(v), is obtained as svt in Standalone Equation 1. The BC of node v is the number of shortest paths from node s to node t (s,t≠v). Finally, the BC should be divided by the possible number of connected nodes, (N-1)(N-2)/2, where N is the number of nodes in the network. If all the nodes go through v in the shortest path, g(v) is equal to 1.

The BC for node b is calculated in Figure 1 and Standalone Equation 2.

**Figure 1.** Calculation of betweenness centrality.




The two nodes (ie, a and e) have two equal shortest paths (ie, abce and abde). The number of shortest paths from node a to node e is 2.

The method used to ensure there are no authors with duplicate names in the network is to identify the large bubble (with high BC) by clicking the linked coauthors and checking if the author is identical between any two neighbor subnetworks (see Multimedia Appendix 1 and 2).

### The Author-Weighted Scheme

The AWS and the author impact factor (AIF) calculations are shown in Standalone Equations 3 and 4:

Considering a paper of m+1 authors with the last being the corresponding author, $W_j$ denotes the weight for an author on the order j in the article byline. The power, $\gamma_j$, is an integer number from m–1 to 0 in descending order. The sum of author weights in a byline is Standalone Equation 5.

The sum of authorships equals 1 for each paper referred to in Standalone Equation 5. This is a basic concept ensuring that all papers have an equal weight irrespective of the number of coauthors [26]. Accordingly, more importance is given to the first (exp[m], primary) and the last (exp[m–1], corresponding or supervisory) authors, whereas it is assumed that the others (the middle authors) have made smaller contributions [27,28]. In Standalone Equation 5, the smallest portion (exp(0)=1) is assigned to the last second author with the odds=1 as the basic reference [29,30].

### Pattern of Author and Nation Collaboration in JMIR mHealth and uHealth

We selected JMIR mHealth and uHealth as the target journal. The authors (n1=3522) (see Multimedia Appendix 3) were collected. The most cited authors using citation analysis were plotted on Google Maps. Bibliometric indices, including the h-, g-, and x-indexes [31-33], the mean of core articles on g(Ag) (citations on g-core/publications on g-core), and the AIF [34,35] for representing individual research achievements were used to evaluate authors and article topics (ie, the keyword clusters). The most highly cited authors can be plotted with a dashboard on Google Maps using the Kano diagram [36,37] to display it. The authors' x-indexes are located on the X-axis, the h-index is on the Y-axis, and the bubbles are sized by AIF and colored by type within four dragrants (ie, from I to IV denoted by the fearure of excellence, citation-oriended, low performance, and production-oriended, respectively). It is worth noting that the Kano diagram separates all authors into three parts (ie, the h-index originated excitement, the one-dimension performance, and the x-index-originated achievement) [36,37].

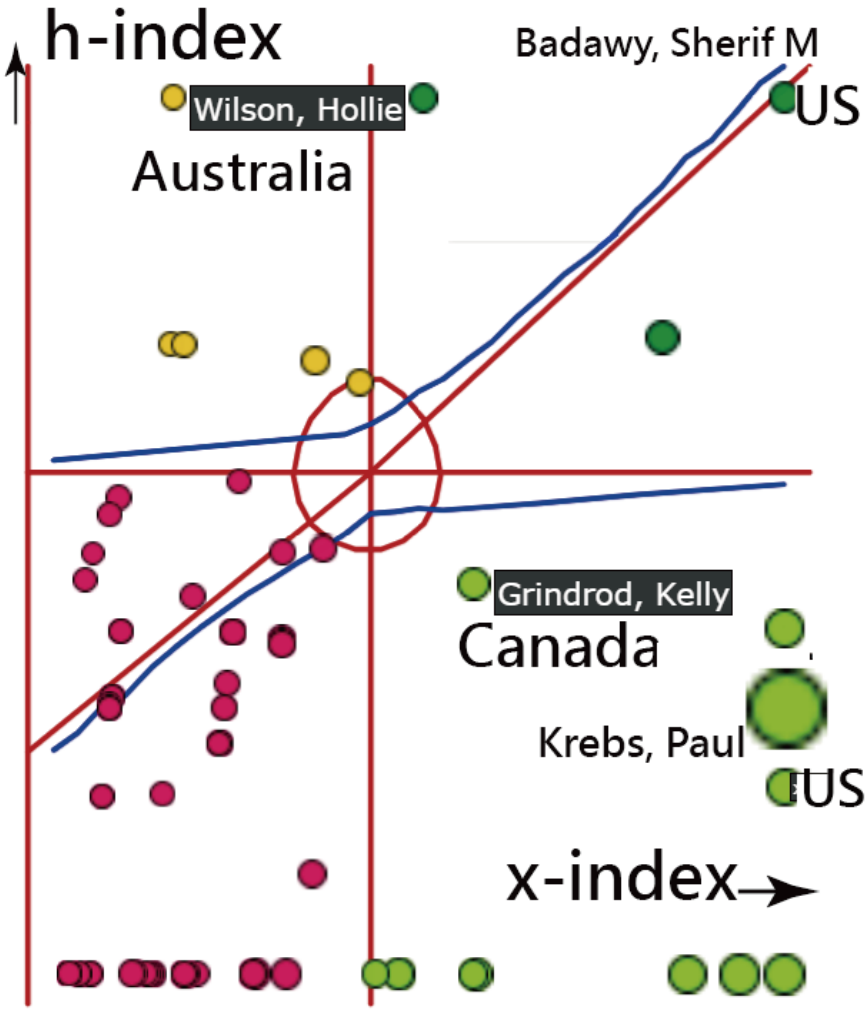
The countries/areas of authors for each published paper were extracted to show the distribution of countries/areas on Google Maps using choropleth maps [38]. The darker regions indicate the most pivotal (or influential) role or bridge in the network if the BC algorithm is performed. Furthermore, the top ten keyword clusters were particularly extracted by SNA, and the representatives with the highest BC in their respective clusters were highlighted on Google Maps. SNA thus filtered the author-defined keywords (n2=1678). Details about the graphical process using SNA and Google Maps are illustrated in Multimedia Appendices 4 and 5.

## Results

### The Most Cited Authors Shown on Google Maps

The most-cited author is Sherif M Badawy (from the United States), who published six articles on JMIR mHealth and uHealth with high bibliometric indices (h=3; AIF=8.47; x=4.68; Ag=5.26). His top five weighted citations are 9.5 ,7.6, 7.3, 1.3, and 0.5, which yield an h-index of 3 at the third position due to the fourth cited value (1.3) being less than the paper number of 4. The Ag (5.26) and x-index (4.68) are yielded because of g being at 5 (ie, the total citations (26.29) are greater than 25) and x at 3 [ci = 7.3 when computing  ], respectively. The biggest bubble denotes the author Paul Krebs from the United States, who has the highest AIF because one of his articles [39] was cited 178 time in the past. Interested authors can scan the QR-code in Figure 2 [40] to examine the various authors' publication outputs and details in PMC by clicking the bubble of a specific author.

**Figure 2.** Authors' citations dispersed on Google Maps.




### Pattern of Countries/Areas Distributed by the First Author

Figure 3 [41] shows the county/area distribution on Google Maps, indicating most "bridge" coauthors are from two countries, the United States and the United Kingdom, using the BC algorithm.

The top six countries with the highest increase in number of production outputs (ie, Growth>0.90) were the United States, the United Kingdom, South Korea, Canada, Australia, and New Zealand (Table 1). The top two countries with the highest proportion of papers produced were the United States (36.83%) and Australia (9.47%). The x-indexes for each country/area are present in the last column in Table 1. It is worth noting that the x-index for JMIR mHealth and uHealth is 26.56, as shown in the bottom right corner.

**Figure 3.** Dispersion of country/area on author collaborations for JMIR mHealth and uHealth.

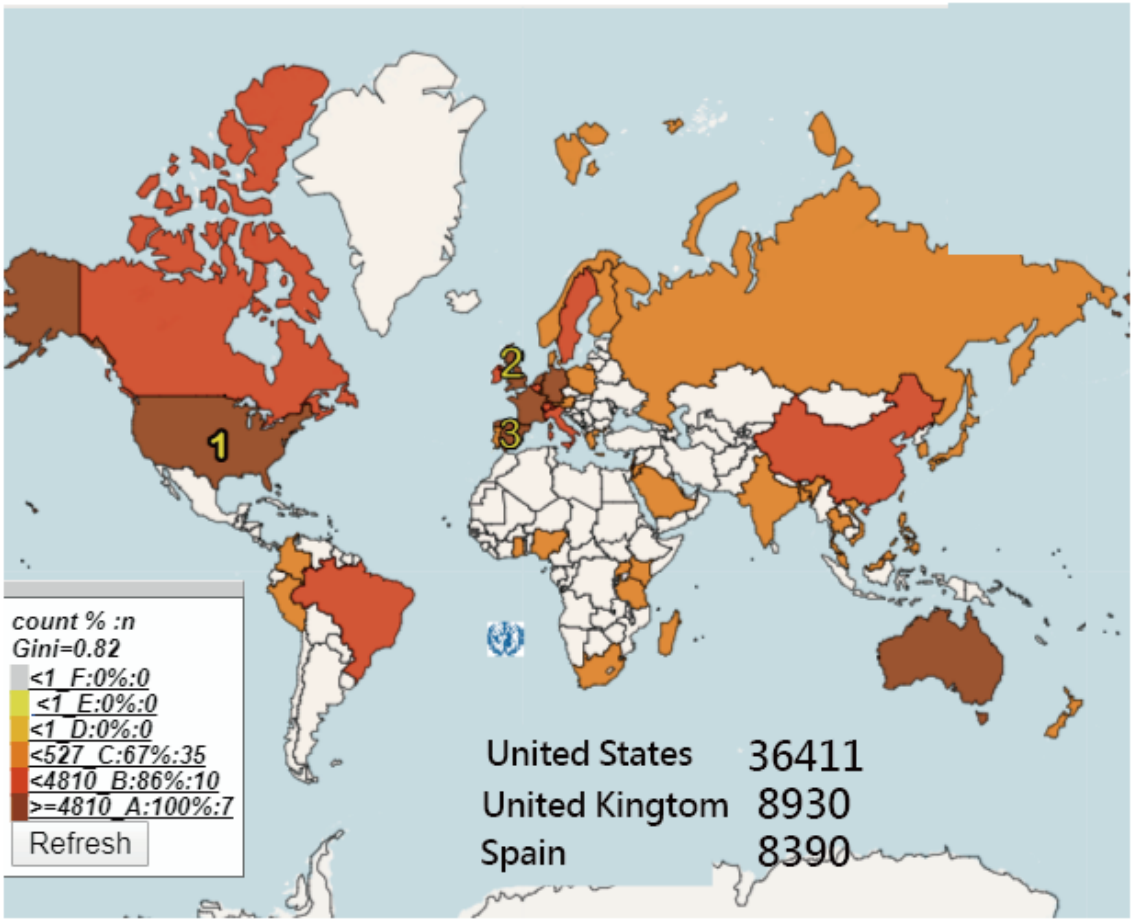


**Table 1.** Dispersions of author collaboration across continents over the years

| Continent, Country | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total, n (%) | Growth[a] | x-index |
|---|---|---|---|---|---|---|---|---|---|
| **Africa** | __[b] | 2 | 1 | 2 | 2 | 1 | 8 (1.18) | 0.71 | — |
| Kenya | — | — | 1 | — | — | — | 1 (0.15) | — | 1.95 |
| Nigeria | — | — | — | — | 1 | — | 1 (0.15) | 0.71 | — |
| South Africa | — | 2 | — | 2 | 1 | — | 5 (0.74) | 0.32 | 2.42 |
| Uganda | — | — | — | — | — | 1 | 1 (0.15) | — | — |
| **Asia** | 3 | 10 | 8 | 9 | 22 | 32 | 84 (12.43) | 0.83 | — |
| China | 2 | 2 | 1 | 1 | 7 | 12 | 25 (3.7) | 0.57 | 3.19 |
| South Korea | — | — | 2 | 2 | 4 | 6 | 14 (2.07) | 0.94 | 3.08 |
| Singapore | — | 3 | — | — | 1 | 4 | 8 (1.18) | –0.12 | 3.56 |
| Thailand | — | 2 | 2 | — | 1 | 2 | 7 (1.04) | — | 2.25 |
| Taiwan | — | — | — | 1 | 2 | 3 | 6 (0.89) | 0.88 | 1.39 |
| Others | 1 | 3 | 3 | 5 | 7 | 5 | 24 (3.55) | 0.97 | — |
| **Europe** | 15 | 12 | 18 | 35 | 60 | 67 | 207 (30.62) | 0.89 | — |
| United Kingdom | 2 | — | 9 | 9 | 13 | 12 | 45 (6.66) | 0.91 | 6.65 |
| Germany | 2 | 2 | 1 | 2 | 11 | 11 | 29 (4.29) | 0.68 | 5.97 |
| Spain | 5 | 1 | 1 | 4 | 5 | 10 | 26 (3.85) | 0.23 | 5.41 |
| Netherlands | 1 | — | 1 | 9 | 7 | 6 | 24 (3.55) | 0.81 | 4.7 |
| Sweden | — | 3 | 4 | 4 | 3 | 4 | 18 (2.66) | 0.67 | 4.84 |
| Others | 5 | 6 | 2 | 7 | 21 | 24 | 65 (9.62) | 0.71 | — |
| **North America** | 6 | 21 | 52 | 70 | 90 | 54 | 293 (43.34) | 0.99 | — |
| United States | 6 | 17 | 42 | 58 | 79 | 47 | 249 (36.83) | 0.99 | 17.13 |
| Canada | — | 4 | 10 | 12 | 11 | 7 | 44 (6.51) | 0.92 | 8.74 |
| **Oceania** | 1 | 9 | 15 | 21 | 19 | 11 | 76 (11.24) | 0.93 | — |
| Australia | 1 | 8 | 13 | 17 | 15 | 10 | 64 (9.47) | 0.91 | 11.03 |
| New Zealand | — | 1 | 2 | 4 | 4 | 1 | 12 (1.78) | 0.97 | 4.81 |
| **South America** | — | 3 | 1 | — | 3 | 1 | 8 (1.18) | 0.31 | — |
| Brazil | — | 2 | — | — | 2 | 1 | 5 (0.74) | 0.29 | 2.52 |
| Colombia | — | 1 | — | — | — | — | 1 (0.15) | –0.35 | 1.59 |
| Peru | — | — | 1 | — | 1 | — | 2 (0.3) | 0.58 | 1.59 |
| Total | 25 | 57 | 95 | 137 | 196 | 166 | 676 (100) | 0.99 | 26.56 |

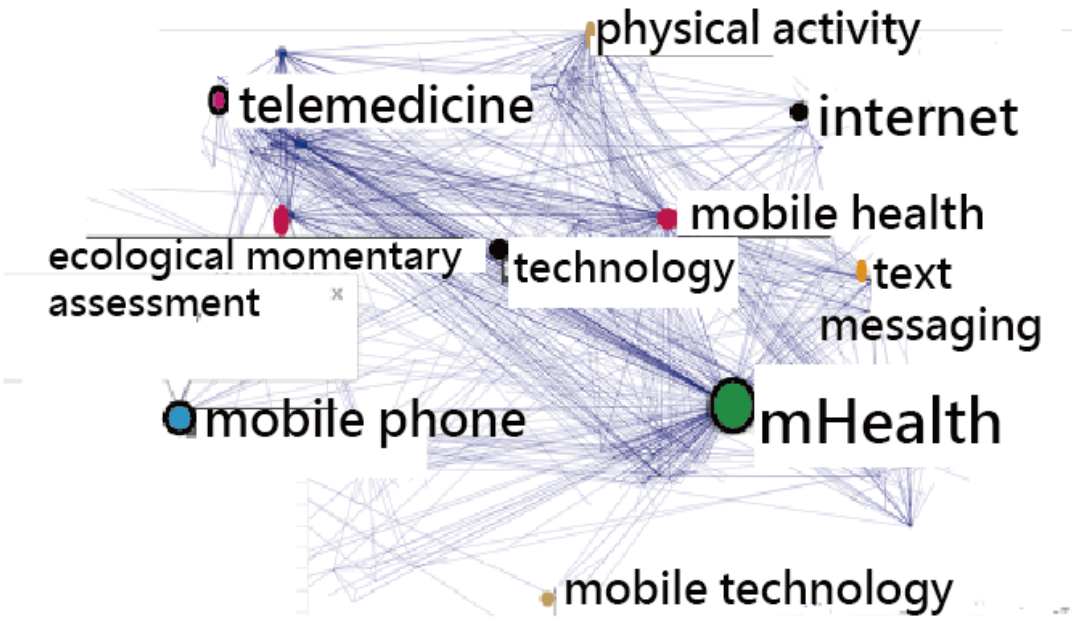
[a]Growth based on data from 2013 and 2017.

[b]Not applicable.

## Clusters of Keywords

The top ten keyword clusters are presented in Figure 4. The representative terms with the highest betweenness centrality are shown for each cluster. The biggest one is that of “mHealth.” It is recommended that interested readers should scan the QR-code in Figure 4 [42] to see the details of the information on Google Maps.

**Figure 4.** Dispersion of keyword clusters for the first author clusters of JMIR mHealth and uHealth. mHealth: mobile health.




## Analyses of Article Topics Related to Bibliometric Indices

The numbers of citable and cited articles across the keyword clusters are shown in Tables 2 and 3. Five bibliometric indices are present at the right-hand side. We found that the AIF had a weak relation with the other four indices, as shown in the bottom right side in Table 2. However, the journal impact factor is 4.37, equivalent to the impact factor of journal citation report (JCR IF)=4.541 in 2017. The two keyword clusters of mHealth and telemedicine earned the highest indices in comparison to their counterparts (Figure 5), indicating both topics have a higher metric (ie, the normalized mean of h, g, x, and Ag) than the other topic clusters.

**Table 2.** Bibliometric indices for medical subject heading (MeSH) terms over the years for publications.

| Keywords | Publication count | | | | | | | AIF[a] | h | g | x | (g)Ag[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2013 (n) | 2014 (n) | 2015 (n) | 2016 (n) | 2017 (n) | 2018 (n) | Total (N) | | | | | |
| Text messaging | —[c] | 4 | 4 | 5 | 6 | 6 | 25 | 4 | 7 | 9 | 7.48 | 9.67 |
| mHealth[d] | 7 | 16 | 39 | 51 | 68 | 55 | 236 | 4.4 | 16 | 21 | 19.13 | 21.57 |
| Physical activity | 2 | 3 | 4 | 8 | 16 | 14 | 47 | 2.83 | 6 | 11 | 7.21 | 11.18 |
| Telemedicine | 2 | 11 | 18 | 33 | 57 | 51 | 172 | 4.87 | 15 | 23 | 16.43 | 24.26 |
| Mobile health | 3 | 8 | 9 | 14 | 21 | 15 | 70 | 4.6 | 10 | 13 | 12.41 | 14.08 |
| Ecological momentary assessment | — | — | 1 | 2 | 2 | 1 | 6 | 1.17 | 1 | 1 | 2.24 | 5 |
| Internet | 3 | 4 | 6 | 3 | 5 | 4 | 25 | 7.36 | 8 | 13 | 9.54 | 14 |
| Obesity | 1 | 2 | 5 | 8 | 4 | 1 | 21 | 5.9 | 6 | 10 | 6.93 | 10.4 |
| Wearable | — | — | 1 | — | 1 | 3 | 5 | 1 | 1 | 1 | 2 | 3 |
| Mobile phone | 1 | 2 | 2 | 6 | 3 | 2 | 16 | 3.56 | 5 | 7 | 5.48 | 7.29 |
| Others | 6 | 7 | 6 | 6 | 13 | 10 | 48 | 2.63 | — | — | — | — |
| Total | 25 | 57 | 95 | 136 | 196 | 162 | 671 | 4.37 | — | — | — | — |

[a]AIF: author impact factor.

[b](g)Ag: publications on g-core.

[c]Not applicable.

[d]mHealth: mobile health.

**Table 3.** Correlation coefficients of metrics for medical subject heading (MeSH) terms over the years for quantity of citations.

| Keywords | Publication count | | | | | | | Correlation | AIF[a] | h | g | x | (g)Ag[b] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2013 (n) | 2014 (n) | 2015 (n) | 2016 (n) | 2017 (n) | 2018 (n) | Total (N) | | | | | | |
| Text messaging | —[c] | 28 | 28 | 30 | 14 | 0 | 100 | AIF | 1 | — | — | — | — |
| mHealth[d] | 112 | 212 | 335 | 242 | 131 | 7 | 1039 | h | 0.57 | 1 | — | — | — |
| Physical activity | 25 | 18 | 19 | 48 | 23 | 0 | 133 | g | 0.63 | 0.98 | 1 | — | — |
| Telemedicine | 46 | 182 | 307 | 186 | 95 | 22 | 838 | x | 0.54 | 0.99 | 0.96 | 1 | — |
| Mobile health | 11 | 82 | 91 | 100 | 38 | 0 | 322 | Ag | 0.58 | 0.98 | 0.99 | 0.96 | 1 |
| Ecological momentary assessment | — | — | 2 | 5 | 0 | 0 | 7 | — | — | — | — | — | — |
| Internet | 33 | 57 | 81 | 9 | 4 | 0 | 184 | — | — | — | — | — | — |
| Obesity | 16 | 12 | 59 | 25 | 12 | 0 | 124 | — | — | — | — | — | — |
| Wearable | — | — | 3 | — | 2 | 0 | 5 | — | — | — | — | — | — |
| Mobile phone | 7 | 10 | 25 | 15 | 0 | 0 | 57 | — | — | — | — | — | — |
| Others | 20 | 35 | 46 | 23 | 2 | 0 | 126 | — | — | — | — | — | — |
| Total | 270 | 636 | 996 | 683 | 321 | 29 | 2935 | — | — | — | — | — | — |

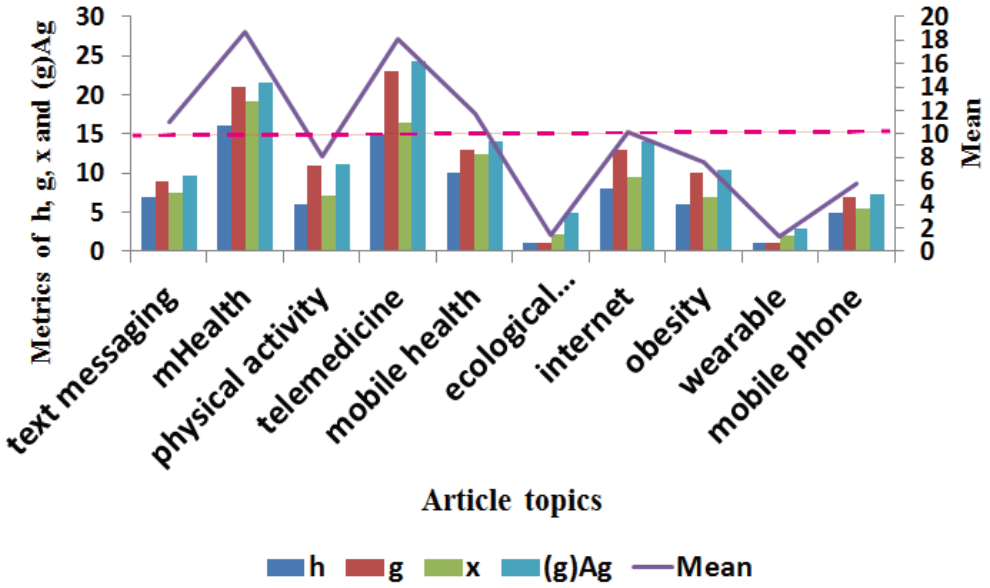
[a]AIF: author impact factor.

[b](g)Ag: publications on g-core.

[c]Not applicable.

[d]mHealth: mobile health.

**Figure 5.** Comparison of article topics related to bibliometric indices. Ag: publication on g-core.




## Discussion

### Principal Findings

We found that the most-cited author is Sherif M Badawy (from the United States), who has published six articles on JMIR mHealth since 2016. Other authors also gained excellent citation indices on Figure 2, such as Stoyan R Stoyanov from the United States (4 papers since 2015), John Torous from Germany (5 papers since 2014), Paul Krebs from Germany (3 papers since 2014), and Kathryn Mercer from Germany (3 papers since 2015). It is easy to examine their publications on PubMed by clicking the author's bubble on Google Maps.

The most productive authors with six papers were Urs-Vito Albrecht (citable=2.6; cited=18.1; AIF=6.8) from Germany, and Sherif M. Badawy (citable=3.3; cited=27.7; AIF=8.5) from the United States. The reason why Badawy has a higher weighted value of citable papers than Albrecht is that the latter was the middle author more often than the former if the AWS in Standalone Equation 3 was applied. If the BCs were applied, the author Ralph Maddison, from Australia, who had five papers

(citable=1.1; cited=6.1; AIF=5.5), played the most pivotal (bridge) role in the authoring network.

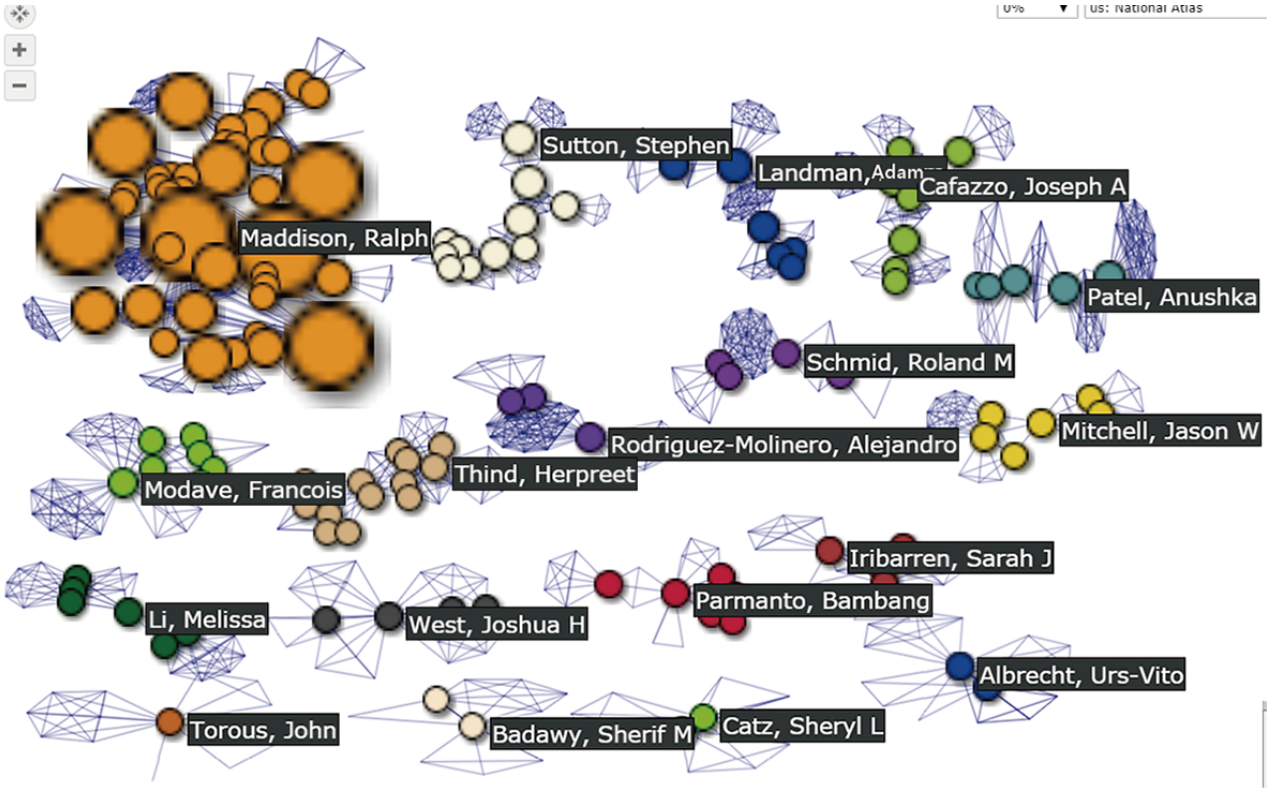
The two countries with the highest BC were the United States (x-index=17.13) and the United Kingdom (x-index=6.65), thereby proving that the United States and Europe still dominate publication output in science [43,44]. Another new finding is about the two keyword clusters of mHealth and telemedicine with the highest metrics among types of article feature, which is rarely seen when combining citation analysis and SNA in previous articles.

## Strength of the Study

Traditionally, in dealing with a test with multiple questions and answers, we often count the item with the highest frequency as representing the most important value. For instance, many customers purchase their goods in a shopping cart, which is like a test of multiple answers without considering any associations between entities. Accordingly, many articles [4-8] merely present the highly frequency counts of authors instead of the association of authors in a network, such as the most productive authors Urs-Vito Albrecht and Sherif M. Badawy in Figure 2, instead of the most pivotal author Ralph Maddison with the highest BC, who is associated with many coauthors in the network. Many data scientists have developed ways to discover new knowledge from the vast quantities of increasingly available information [45], especially by applying SNA [4-6] to large data analysis.

We also ensured that no author had duplicate names in the network via identification of the large bubble (ie, with a high BC) first by clicking the linked coauthors (eg, Francois Modave at the left-bottom bubble in Figure 6), and then checking the author without duplicate names in the network by clicking the associated coauthors in the opposite neighbor subnetworks to examine whether the author had the same names in each paper. The dashboard [46] could easily be linked to the published papers in Medline if the author was clicked. For further details about the steps made to ensure there were no authors without duplicate names, see Multimedia Appendices 1 and 2.

**Figure 6.** Author clusters in a collaboration network.




Furthermore, we found 335 papers in Medline because of the keyword social network analysis (Title) as of May 20, 2018. In practice, we found studies on duplicative prescriptions using SNA in Japan [47] and one explaining HIV risk multiplexity [48]. However, no such study like ours has incorporated the SNA analysis with Google Maps to interpret the results. Many papers investigated most-cited articles or most productive authors in academics. Few inspected most-cited authors in a given journal. Overall, two challenges we faced have been overcome in this study: (1) some different authors with the same name in bibliometric data; and (2) coauthors' contributions differing in the article byline. Furthermore, we illustrated a way to examine article topics associated with the number of citations for a journal.

Previous studies [49-51] reported: (1) a higher impact factor being associated with the publication of reviews and original articles instead of case reports; (2) rigorous systematic reviews receiving more citations than other narrative reviews; and (3) case reports with low impact factors due to them being rarely cited by articles. In comparison, we applied the author-defined keywords to cluster article features, which is different from previous studies in that an objective verification was made for a given journal. As such, the bibliometric metrics can be linked

to the article features if each article has been assigned to its corresponding type.

Regarding the incorporation of Google Maps with SNA, Google Maps are sophisticatedly linked in references [41-52] for readers interested in manipulating the link as a dashboard. The country/area distribution in Figure 3 easily illustrates the feature of international author collaborations in JMIR mHealth and uHealth. We hope subsequent studies can report other types of information using the Google application programming interface to readers in the future.

### Limitations and Future Study

Although findings were based on the above analysis, the results should be interpreted with caution because of several potential limitations. First, this study only focused on a single journal. Any generalization should be made in similar fields of journal contents. Second, although SNA is quite useful in exploring the topic evolution and identifying hotspots for keywords, the results might be affected by the accuracy of the author-defined terms. The medical subject heading (MeSH) terms included in the PubMed library are recommended for use in the future. Third, many different algorithms are used for SNA. We merely applied community cluster and density with BC in the figures. Any changes made along with the algorithm will present different patterns and inferences. Fourth, SNA is not subject to the Pajek software we used in this study. Others, such as Ucinet [53] and Gephi [54], are suggested to readers for use in the future. Fifth, we downloaded citing articles from PMC, which are different from many citation analyses that use other academic databases, such as the Scientific Citation Index, Scopus, and Google Scholar [55-58], to investigate the most cited articles in a specific discipline. This approach using data from PMC can lead to more citation studies reporting the most cited authors in other disciplines.

### Conclusions

The most cited authors were selected using the authorship-weighted scheme (AWS). The keywords of mHealth and telemedicine are potentially highly cited more than other types of keywords. The results on Google Maps are novel and unique as a knowledge concept maps for understanding the features of a journal. The research approaches used in this study (ie, BC and AWS) can be applied to other bibliometric analyses in the future.

### Authors' Contributions

WC conceived and designed the study. WC and TW performed the statistical analyses and were in charge of dealing with data. YT and WC helped design the study, collected information, and interpreted data. PH monitored the research. All authors read and approved the final article.

### Conflicts of Interest

None declared.

Multimedia Appendix 1
MP4: Identifying the unique author name.
[TXT File , 0 KB - mhealth_v8i5e11567_app1.txt ]

Multimedia Appendix 2
PDF:using between centrality to detect authors with duplicate names in a network.
[PDF File (Adobe PDF File), 1583 KB - mhealth_v8i5e11567_app2.pdf ]

Multimedia Appendix 3
Txt:Pajek control file and dataset.
[TXT File , 233 KB - mhealth_v8i5e11567_app3.txt ]

Multimedia Appendix 4
MP4”How to deal with data and build the Google maps.
[TXT File , 0 KB - mhealth_v8i5e11567_app4.txt ]

Multimedia Appendix 5
MP4: MS Excel module extracting data from a website and plotting Google Maps.
[TXT File , 0 KB - mhealth_v8i5e11567_app5.txt ]

### References

1. Avula J, Avula H. Authors, authorship order, the moving finger writes. J Indian Soc Periodontol 2015;19(3):258-262 [FREE Full text] [doi: 10.4103/0972-124X.145782] [Medline: 26229263]

2. Baber Z, Gibbons M, Limoges C, Nowotny H, Schwartzman S, Scott P, et al. The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies. Contemporary Sociology 1995 Nov;24(6):751 [FREE Full text] [doi: 10.2307/2076669]
3. Wu JC, Chiang T. Comparing child mortality in Taiwan and selected industrialized countries. J Formos Med Assoc 2007 Mar;106(2):177-180 [FREE Full text] [doi: 10.1016/S0929-6646(09)60237-0] [Medline: 17339165]
4. Sadoughi F, Valinejadi A, Shirazi M, Khademi R. Social Network Analysis of Iranian Researchers on Medical Parasitology: A 41 Year Co-Authorship Survey. Iran J Parasitol 2016;11(2):204-212 [FREE Full text] [Medline: 28096854]
5. Osareh F, Khademi R, Rostami MK, Shirazi MS. Co-authorship Network Structure Analysis of Iranian Researchers' scientific outputs from 1991 to 2013 based on the Social Science Citation Index (SSCI). Collnet Journal of Scientometrics and Information Management 2015 Feb 10;8(2):263-271 [FREE Full text] [doi: 10.1080/09737766.2014.1015301]
6. Liu X, Bollen J, Nelson ML, Van de Sompel H. Co-authorship networks in the digital library research community. Information Processing & Management 2005 Dec;41(6):1462-1480 [FREE Full text] [doi: 10.1016/j.ipm.2005.03.012]
7. Shen L, Xiong B, Li W, Lan F, Evans R, Zhang W. Visualizing Collaboration Characteristics and Topic Burst on International Mobile Health Research: Bibliometric Analysis. JMIR Mhealth Uhealth 2018 Jul 05;6(6):e135 [FREE Full text] [doi: 10.2196/mhealth.9581] [Medline: 29871851]
8. Chien T, Chang Y, Wang H. Understanding the productive author who published papers in medicine using National Health Insurance Database. Medicine 2018;97(8):e9967. [doi: 10.1097/md.0000000000009967] [Medline: 29465594]
9. Freeman LC. Centrality in social networks conceptual clarification. Social Networks 1978 Jan;1(3):215-239 [FREE Full text] [doi: 10.1016/0378-8733(78)90021-7]
10. Otte E, Rousseau R. Social network analysis: a powerful strategy, also for the information sciences. Journal of Information Science 2016 Jul;28(6):441-453 [FREE Full text] [doi: 10.1177/016555150202800601]
11. Sekercioglu CH. Quantifying coauthor contributions. Science 2008 Oct 17;322(5900):371. [doi: 10.1126/science.322.5900.371a] [Medline: 18927373]
12. Batista PD, Campiteli MG, Kinouchi O. Is it possible to compare researchers with different scientific interests? Scientometrics 2006 Jul;68(1):179-189 [FREE Full text] [doi: 10.1007/s11192-006-0090-4]
13. Lindsey D. Further evidence for adjusting for multiple authorship. Scientometrics 1982 Sep;4(5):389-395 [FREE Full text] [doi: 10.1007/bf02135124]
14. Pu Q, Lyu Q, Liu H, Fan K. Bibliometric analysis of the top-cited articles on islet transplantation. Medicine (Baltimore) 2017 Dec;96(44):e8247 [FREE Full text] [doi: 10.1097/MD.0000000000008247] [Medline: 29095254]
15. Tian J, Li M, Lian F, Tong X. The hundred most-cited publications in microbiota of diabetes research: A bibliometric analysis. Medicine (Baltimore) 2017 Oct;96(37):e7338 [FREE Full text] [doi: 10.1097/MD.0000000000007338] [Medline: 28906350]
16. Miao Y, Liu R, Pu Y, Yin L. Trends in esophageal and esophagogastric junction cancer research from 2007 to 2016: A bibliometric analysis. Medicine (Baltimore) 2017 May;96(20):e6924 [FREE Full text] [doi: 10.1097/MD.0000000000006924] [Medline: 28514311]
17. Zhang Y, Huang J, Du L. The top-cited systematic reviews/meta-analyses in tuberculosis research: A PRISMA-compliant systematic literature review and bibliometric analysis. Medicine (Baltimore) 2017 Feb;96(6):e4822 [FREE Full text] [doi: 10.1097/MD.0000000000004822] [Medline: 28178120]
18. Liao J, Wang J, Liu Y, Li J, Duan L, Chen G, et al. Modern researches on Blood Stasis syndrome 1989-2015: A bibliometric analysis. Medicine (Baltimore) 2016 Dec;95(49):e5533 [FREE Full text] [doi: 10.1097/MD.0000000000005533] [Medline: 27930547]
19. Li H, Zhao X, Zheng P, Hu M, Lu Y, Jia F, et al. Classic Citations in Main Primary Health Care Journals: A PRISMA-Compliant Systematic Literature Review and Bibliometric Analysis. Medicine (Baltimore) 2015 Dec;94(49):e2219 [FREE Full text] [doi: 10.1097/MD.0000000000002219] [Medline: 26656360]
20. Dasgupta S, Vaughan AS, Kramer MR, Sanchez TH, Sullivan PS. Use of a Google Map Tool Embedded in an Internet Survey Instrument: Is it a Valid and Reliable Alternative to Geocoded Address Data? JMIR Res Protoc 2014 Apr 10;3(2):e24 [FREE Full text] [doi: 10.2196/resprot.2946] [Medline: 24726954]
21. Kobayashi S, Fujioka T, Tanaka Y, Inoue M, Niho Y, Miyoshi A. A geographical information system using the Google Map API for guidance to referral hospitals. J Med Syst 2010 Dec;34(6):1157-1160. [doi: 10.1007/s10916-009-9335-0] [Medline: 20703591]
22. Bright CF, Haynes EE, Patterson D, Pisu M. The Value of Social Network Analysis for Evaluating Academic-Community Partnerships and Collaborations for Social Determinants of Health Research. Ethn Dis 2017 Nov 09;27(Suppl 1):337 [FREE Full text] [doi: 10.18865/ed.27.s1.337]
23. deNooy W, Mrvar A, Batagelj V. Exploratory Social Network Analysis With Pajek: Revised and Expanded,2nd edn. New York, NY: Cambridge University Press 2011 [FREE Full text] [doi: 10.1017/cbo9780511996368]
24. Phan TG, Beare R, Chen J, Clissold B, Ly J, Singhal S, et al. Googling Service Boundaries for Endovascular Clot Retrieval Hub Hospitals in a Metropolitan Setting: Proof-of-Concept Study. Stroke 2017 May;48(5):1353-1361. [doi: 10.1161/STROKEAHA.116.015323] [Medline: 28356438]

25. Brandes U. A faster algorithm for betweenness centrality. The Journal of Mathematical Sociology 2001 Jun;25(2):163-177 [FREE Full text] [doi: 10.1080/0022250x.2001.9990249]
26. Vavryčuk V. Fair ranking of researchers and research teams. PLoS ONE 2018 Apr 5;13(4):e0195509 [FREE Full text] [doi: 10.1371/journal.pone.0195509]
27. Egghe L, Rousseau R, Van Hooydonk G. Methods for accrediting publications to authors or countries: Consequences for evaluation studies. J. Am. Soc. Info. Sci 2000;51(2):145-157 [FREE Full text] [doi: 10.1002/(sici)1097-4571(2000)51:2<145::aid-asi6>3.0.co;2-9]
28. Batagelj V, Mrvar A. Pajek - Analysis and Visualization of Large Networks. In: International Symposium on Graph Drawing. Berlin, Germany: Springer Nature; 2003:477-478.
29. Chien T, Chow JC, Chang Y, Chou W. Applying Gini coefficient to evaluate the author research domains associated with the ordering of author names: A bibliometric study. Medicine (Baltimore) 2018 Sep;97(39):e12418 [FREE Full text] [doi: 10.1097/MD.0000000000012418] [Medline: 30278518]
30. Chien T, Wang H, Chang Y, Kan W. Using Google Maps to display the pattern of coauthor collaborations on the topic of schizophrenia: A systematic review between 1937 and 2017. Schizophr Res 2019 Feb;204:206-213. [doi: 10.1016/j.schres.2018.09.015] [Medline: 30262255]
31. Hirsch JE. An index to quantify an individual's scientific research output. Proc Natl Acad Sci U S A 2005 Nov 15;102(46):16569-16572 [FREE Full text] [doi: 10.1073/pnas.0507655102] [Medline: 16275915]
32. Egghe L. Theory and practise of the g-index. Scientometrics 2013 Jun 20;69(1):131-152 [FREE Full text] [doi: 10.1007/s11192-006-0144-7]
33. Fenner T, Harris M, Levene M, Bar-Ilan J. A novel bibliometric index with a simple geometric interpretation. PLoS ONE 2018 Jul 10;13(7):e0200098 [FREE Full text] [doi: 10.1371/journal.pone.0200098]
34. Lippi G, Mattiuzzi C. Scientist impact factor (SIF): a new metric for improving scientists' evaluation? Ann. Transl. Med 2017 Aug;5(15):303-303 [FREE Full text] [doi: 10.21037/atm.2017.06.24]
35. Pan RK, Fortunato S. Author Impact Factor: tracking the dynamics of individual scientific impact. Sci Rep 2014 May 12;4(1) [FREE Full text] [doi: 10.1038/srep04880]
36. Kano N, Seraku N, Takahashi F, Tsuji S. Attractive Quality and Must-Be Quality. Journal of the Japanese Society for Quality Control 1984;41:39-48 [FREE Full text]
37. Lin C, Chou P, Chou W, Chien T. Using the Kano model to display the most cited authors and affiliated countries in schizophrenia research. Schizophrenia Research 2019 Dec. [doi: 10.1016/j.schres.2019.10.058] [Medline: 31862218]
38. Chien T, Wang H, Hsu C, Kuo S. Choropleth map legend design for visualizing the most influential areas in article citation disparities: A bibliometric study. Medicine (Baltimore) 2019 Oct;98(41):e17527 [FREE Full text] [doi: 10.1097/MD.0000000000017527] [Medline: 31593127]
39. Krebs P, Duncan DT. Health App Use Among US Mobile Phone Owners: A National Survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
40. Chien T. The most-cited authors who published papers in JMIR mHealth and uHealth using authorship-weighted scheme: A Bibliometric Analysis. JMIR Preprints 2018 Jul 6:1-29 [FREE Full text] [doi: 10.2196/11567]
41. Chien T. The nation dispersion on Google Maps for JMIR mHealth and uHealth. 2020 Feb 1. URL: http://www.healthup.org.tw/gps/JMIRmhealthauaif.htm [accessed 2020-02-01]
42. Chien T. The nation dispersion on Google Maps for JMIR mHealth and uHealth. 2018 Feb 2. URL: http://www.healthup.org.tw/gps/JMIRmhealthnation.htm [accessed 2020-02-02]
43. Leydesdorff L, Wagner CS, Park H, Adams J. International collaboration in science: the global map and the network. El Profesional de la Informacion 2013 Feb 12;22(1):87-95 [FREE Full text] [doi: 10.3145/epi.2013.ene.12]
44. Glänzel W, Schlemmer B. National research profiles in a changing Europe (1983–2003) An exploratory study of sectoral characteristics in the Triple Helix. Scientometrics 2007 Feb;70(2):267-275 [FREE Full text] [doi: 10.1007/s11192-007-0203-8]
45. Verhoef P, Kooge E, Walk N. Creating Value with Big Data Analytics: Making Smarter Marketing Decisions. London: Routledge 2016 [FREE Full text] [doi: 10.4324/9781315734750]
46. Chien T. The author network on Google Maps for JMIR mHealth and uHealth. 2020 Feb 1. URL: http://www.healthup.org.tw/gps/JMIRmhealthego.htm [accessed 2020-02-01]
47. Takahashi Y, Ishizaki T, Nakayama T, Kawachi I. Social network analysis of duplicative prescriptions: One-month analysis of medical facilities in Japan. Health Policy 2016 Mar;120(3):334-341 [FREE Full text] [doi: 10.1016/j.healthpol.2016.01.020]
48. Felsher M, Koku E. Explaining HIV Risk Multiplexity: A Social Network Analysis. AIDS Behav 2018 Nov 21;22(11):3500-3507. [doi: 10.1007/s10461-018-2120-7] [Medline: 29680933]
49. Rodríguez-Lago L, Molina-Leyva A, Pereiro-Ferreirós M, García-Doval I. Influence of Article Type on the Impact Factor of Dermatology Journals. Actas Dermo-Sifiliográficas (English Edition) 2018 Jun;109(5):432-438. [doi: 10.1016/j.adengl.2018.04.002] [Medline: 29496199]
50. Bhandari M, Montori VM, Devereaux PJ, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Doubling the impact: publication of systematic review articles in orthopaedic journals. J Bone Joint Surg Am 2004 May;86(5):1012-1016. [Medline: 15118046]

51. Nielsen M, Seitz K. Impact Factors and Prediction of Popular Topics in a Journal. Ultraschall Med 2016 Aug;37(4):343-345. [doi: 10.1055/s-0042-111209] [Medline: 27490462]
52. Chien T. The keyword dispersion on Google Maps for JMIR mHealth and uHealth. 2018 Feb 1. URL: http://www.healthup.org.tw/gps/JMIRmobolkey.htm [accessed 2020-02-01]
53. Borgatti S, Everett M, Freeman L. Ucinet for Window: software for Social Network Analysis. In: Encyclopedia of Social Network Analysis and Mining. New York City, New York: Springer; Oct 5, 2014:2002.
54. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. 2020 Feb 1 Presented at: International AAAI Conference on Weblogs and Social Media; May 17-20; San Jose, California URL: https://gephi.org/ [doi: 10.1007/978-1-4614-7163-9_299-1]
55. Alotaibi NM, Nassiri F, Badhiwala JH, Witiw CD, Ibrahim GM, Macdonald RL, et al. The Most Cited Works in Aneurysmal Subarachnoid Hemorrhage: A Bibliometric Analysis of the 100 Most Cited Articles. World Neurosurg 2016 May;89:587-592.e6. [doi: 10.1016/j.wneu.2015.11.072] [Medline: 26723285]
56. Thulesius H. Assessing research impact with Google Scholar: The most cited articles in the journal 2008–2010. Scandinavian Journal of Primary Health Care 2011 Nov 29;29(4):193-195 [FREE Full text] [doi: 10.3109/02813432.2011.629160]
57. Nuti SV, Ranasinghe I, Murugiah K, Shojaee A, Li S, Krumholz HM. Association Between Journal Citation Distribution and Impact Factor. Journal of the American College of Cardiology 2015 Apr;65(16):1711-1712 [FREE Full text] [doi: 10.1016/j.jacc.2014.12.071]
58. Gini C. Variability and mutability, contribution to the study of statistical distribution and relations. Studi Economico-Giuricici della R. Universita de Cagliari 1912 [FREE Full text]


## Abbreviations

**AIF:** author impact factor
**AWS:** authorship-weighted scheme
**BC:** betweenness centrality
**g(Ag):** citations on g-core/publications on g-core
**MeSH:** medical subject heading
**PMC:** PubMed Central
**SNA:** social network analysis
**VBA:** visual basic for applications


*Edited by G Eysenbach; submitted 11.07.18; peer-reviewed by F Soheili, L Shen; comments to author 08.10.18; revised version received 22.10.18; accepted 26.01.20; published 07.05.20.*

*Please cite as:*
*Kan WC, Chou W, Chien TW, Yeh YT, Chou PH*
*The Most-Cited Authors Who Published Papers in JMIR mHealth and uHealth Using the Authorship-Weighted Scheme: Bibliometric Analysis*
*JMIR Mhealth Uhealth 2020;8(5):e11567*
*URL: https://mhealth.jmir.org/2020/5/e11567*
*doi:10.2196/11567*
*PMID:32379053*

Original Paper

# Evaluating the Quality of Health-Related WeChat Public Accounts: Cross-Sectional Study

Fuzhi Wang[1,2,3*], PhD; Zhuoxin Wang[2,4*], MSc; Weiwei Sun[1,2*], MSc; Xiumu Yang[5*], PhD; Zhiwei Bian[1,2*], PhD; Lining Shen[2,6*], PhD; Wei Pan[1,2*], PhD; Peng Liu[1,2*], MSc; Xingzhi Chen[1,2*], MSc; Lianguo Fu[2,7*], PhD; Fan Zhang[1*], BSc; Dan Luo[1,2,5*], MSc

[1]School of Health Management, Bengbu Medical College, Bengbu, China

[2]Innovation Team of Health Information Management and Application Research (BYKC201913), Bengbu Medical College, Bengbu, China

[3]Project Team of Outstanding Young Teachers, Bengbu Medical College, Bengbu, China

[4]School of Nursing, Bengbu Medical College, Bengbu, China

[5]The General Practice Medical Education and Development Center of Anhui Province, Bengbu Medical College, Bengbu, China

[6]School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, China

[7]School of Public Health, Bengbu Medical College, Bengbu, China

[*]all authors contributed equally

**Corresponding Author:**
Dan Luo, MSc
School of Health Management
Bengbu Medical College
2600 Donghai Rd, Longzihu District
Bengbu
China
Phone: 86 13865073216
Email: ld.bbmc@foxmail.com

## Abstract

**Background:** As representatives of health information communication platforms accessed through mobile phones and mobile terminals, health-related WeChat public accounts (HWPAs) have a large consumer base in the Chinese-speaking world. However, there is still a lack of general understanding of the status quo of HWPAs and the quality of the articles they release.

**Objective:** The aims of this study were to assess the conformity of HWPAs to the Health on the Net Foundation Code of Conduct (HONcode) and to evaluate the suitability of articles disseminated by HWPAs.

**Methods:** The survey was conducted from April 23 to May 5, 2019. Based on the monthly (March 1-31, 2019) WeChat Index provided by Qingbo Big Data, the top 100 HWPAs were examined to evaluate their HONcode compliance. The first four articles published by each HWPA on the survey dates were selected as samples to evaluate their suitability. All materials were assessed by three raters. The materials were assessed using the HONcode checklist and the Suitability Assessment of Materials (SAM) score sheet. Data analysis was performed with SPSS version 17.0 (SPSS Inc, Chicago, IL, USA) and Excel version 2013 (Microsoft Inc, Washington DC, USA).

**Results:** A total of 93 HWPAs and 210 of their released articles were included in this study. For six of the eight principles, the 93 HWPAs nearly consistently did not meet the requirements of the HONcode. The HWPAs certified by Tencent Corporation (66/93, 71%) were generally slightly superior to those without such certification (27/93, 29%) in terms of compliance with HONcode principles. The mean SAM score for the 210 articles was 67.72 (SD 10.930), which indicated “adequate” suitability. There was no significant difference between the SAM scores of the articles published by certified and uncertified HWPAs ($P$=.07), except in the literacy requirements dimension ($t_{df=97}$=–2.418, $P$=.02).

**Conclusions:** The HWPAs had low HONcode conformity. Although the suitability of health information released by HWPAs was at a moderate level, there were still problems identified, such as difficulty in tracing information sources, excessive implicit advertisements, and irregular usage of charts. In addition, the low approval requirements of HWPAs were not conducive to improvement of their service quality.

***(JMIR Mhealth Uhealth 2020;8(5):e14826)*** doi:10.2196/14826

**KEYWORDS**

health-related WeChat Public Account; HONcode; suitability assessment of material; evaluation; social media; mHealth; app; health information; internet

## Introduction

With progress of information technology, internet-based new media application platforms are becoming an important resource for the public to obtain health information [1,2]. Compared with traditional health information sources, internet-based online health information (OHI) is widely distributed, abundant, rapidly growing, and diverse in form. A common view about the quality of OHI is that a substantial amount of health information should be produced in a specific medical context, which is often ignored in the internet age, leading to members of the public taking OHI out of context. In addition, many websites often provide links to network information with irrelevant or even fake content for commercial benefit [3]. Therefore, evaluation of the quality of OHI has attracted worldwide attention.

A study conducted at China Renmin University investigated the quality of Chinese health information services with a self-constructed evaluation system and found that privacy protection, information integrity, accessibility, and the platform response of information services were important factors influencing the improvement of user participation [4]. Another study performed at Wuhan University constructed an OHI quality evaluation standard system consisting of two primary indicators, seven secondary indicators, and seven tertiary indicators. In addition, suggestions were proposed for the construction of health information websites and the improvement of the quality of network health information from the perspective of users [5]. These articles used self-compiled evaluation tools to study the quality of OHI, whereas international evaluations of the quality of OHI have mostly used more mature evaluation tools. A review of 70 studies on the quality of OHI in 18 countries showed that the most commonly used evaluation tools globally include DISCERN [6], Health on the Net Foundation Code of Conduct (HONcode) [7], Journal of American Medical Association benchmark [8], and LIDA instrument [9]. Approximately 50% of the 70 studies resulted in completely negative evaluations of OHI, and 27.1% of the studies had both positive and negative evaluations. The main reasons for a negative evaluation included the website organizer and sponsor information were not transparent, the disease description and drug information were not accurate, and the information source and authors' identities were not disclosed [10]. Another comparative analysis of OHI services in China and the United States showed that the certification of OHI services in China mainly involves official website certification and internet drug information service certification, whereas the United States mainly focuses on HONcode (health information quality certification) and TRUSTe (website safety certification). Although the content of the different website certification systems is similar, the Chinese health website certification system is more focused on evaluating external features such as website structure and services but lacks evaluation of OHI quality [11].

Compared with traditional health information dissemination through a website, WeChat-based health information dissemination is more convenient. Users can "send out" health information to a broad community through WeChat groups through a simple sharing operation via mobile terminals. Therefore, as a new medium of internet information dissemination, WeChat has a unique impact on the dissemination of Chinese OHI.

As online information communication platforms launched by Tencent, WeChat public accounts (WPAs) have been popular in Chinese-speaking communities worldwide. According to the WeChat Data Report 2018 released by Tencent in 2018, the monthly number of active WPAs exceeded 3.5 million, and the monthly number of active fans reached 797 million [12]. WeChat has gained rapid popularity in mainland China, Taiwan, Hong Kong, Macao, and other regions in the world where people of Chinese ethnicity reside, so that WPAs are now an important means of disseminating information [13]. Unlike apps on mobile platforms, WPAs do not differ based on the operating system; both Android and iOS support access to WPAs. In addition, WPAs feature timely information push notifications, content relevant to everyday life, along with light and humorous writing [14].

Health-related WeChat public accounts (HWPAs) are in a stage of rapid development [15]. By April 22, 2019, the top 100 HWPAs according to the WeChat Communication Index (WCI) had published more than 11,000 articles in total, with a total article access count of over 247 million (see Multimedia Appendix 1, downloaded 9:22 am April 22, 2019). HWPAs have an important impact on public health education and health promotion. However, the number of HWPAs is large. From a content perspective, there are official WPAs from medical institutions as well as information service public accounts dedicated to health care, disease rehabilitation, and other health knowledge dissemination. WPA owners include companies, government agencies, nonprofit organizations, and individuals. Although it is convenient for the public to obtain health-related information from WPAs, there is increasing doubt about the quality of the health information released through WPAs [16]. To ensure the authenticity and security of WPAs, the Tencent Corporation provides an authentication service for WPAs. For verified WPAs, the authentication information and WeChat authentication unique identity are displayed in the authentication details of the account. However, certification is not mandatory. Individuals and organizations can still apply for WPAs and release health-related information even without official Tencent certification.

In the past 5 years, a large number of studies have examined the use of WPAs in the fields of health education [17] and health intervention [18,19]; however, there is still a lack of general understanding about the quality of HWPAs and the articles released on such platforms. Accordingly, the aim of this study was to evaluate the HONcode conformity of HWPAs and

analyze the suitability of articles posted by HWPAs to provide support for improving the service quality of HWPAs and optimizing the OHI communication environment.

## Methods

### Sample Selection

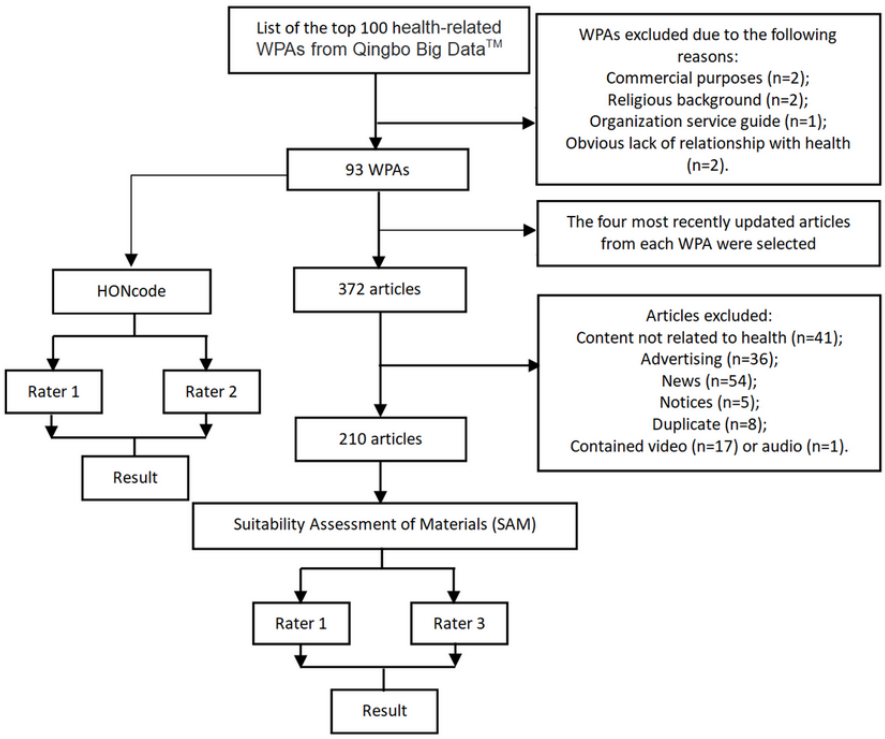
Many organizations and companies have proposed their own evaluation standards to evaluate the influence of WPAs. One of the most widely used standards is the WCI proposed by Qingbo Bigdata Technology Co Ltd (Beijing, China). Qingbo Big Data is well-known among researchers and policymakers of the new media influence evaluation criterion in China by providing big data technology services for the media, public opinion and industry, and their customers, including the Chinese government, top Chinese news media (eg, Xinhua News Agency, People's Daily, China National Radio), and large multinational enterprises [20]. The WCI consists of four primary indicators (spread rate of the whole article, average spread rate of each article, title spread rate, and peak spread rate), eight secondary indicators, and a set of calculation formulas for standardized scores [21]. A higher WCI value represents a larger WPA influence. The latest version of the WCI is version 13.0, updated in January 2017.

We searched the health category of the WPA monthly list (March 1-31, 2019) provided by Qingbo Big Data. The first 100 HWPAs in the WCI were selected as the survey sample. The exclusion criteria for HWPAs were as follows: (1) commodity sales as the main purpose, (2) religious background, (3) organization service guide, and (4) obvious lack of relationship with health. Finally, 93 HWPAs were included in this study. Some examples of the HWPAs are provided in Figure 1.

**Figure 1.** Examples of WeChat public accounts (WPAs). Left: Search results for health-related WPAs (HWPAs) from the keyword 健康 (Health) in the WeChat app. Middle: Client home page of a HWPA. Clicking 关注 (Follow) allows the user to follow the articles published by the HWPA, and clicking the icon and text below allows the user to read the articles published by the HWPA. Right: The HWPA menu bar. The icon below provides information classification support and interaction support for users. Retrieval date: April 22, 2019.


The article sample pool was formed by taking four articles that were newly released by each HWPA on the survey dates (April 23 to May 5, 2019) for a total of 372 articles. The exclusion criteria for articles were as follows: (1) duplicate articles, (2) content not related to health knowledge, (3) advertising/news/notices, and (4) video or audio materials. Finally, 210 articles met the inclusion criteria. Figure 2 illustrates the search and screening flow for the HWPAs and articles.

**Figure 2.** Search and screening flow for health-related WeChat public accounts (HWPAs) and articles. HONcode: Health on the Net Foundation Code of Conduct.




## Evaluation Tools

To the best of our knowledge, there is no quality assessment tool for WPAs, which are online information dissemination platforms that are similar to websites in terms of information release, content services, and operation modes. Therefore, in this study, the HONcode scale was used as the tool to evaluate HWPA quality specifications. Health on the Net (HON) is an international nongovernmental and nonprofit organization established in Switzerland in 1996. The HONcode for medical and health websites addresses one of the internet's main health care issues: the reliability and credibility of information. HONcode provides a set of basic ethical standards for website developers to adhere to with respect to the presentation of information, and aims to ensure that readers always know the source and purpose of the data they are reading [7]; it is currently the most widely used code of ethics for the quality of OHI in the world [22]. As online information platforms used to disseminate health knowledge to the public, HWPAs should also follow the HONcode in the construction of their platforms and the determination of their behaviors in OHI dissemination. Therefore, we believe that it is necessary and appropriate to evaluate HONcode compliance for HWPAs.

Although there are many tools available for assessing the quality of OHI, such as the mHONcode, Michigan website evaluation checklist, LIDA scales, DISCERN instrument, and Simple Measure of Gobbledygook (SMOG) readability formula, none of these tools is well suited for assessing the quality of health information on mobile information platforms. The mHONcode is suitable for evaluating health-related apps that run independently on Android or iOS [23], while WPAs are a platform established on the WeChat app. The Michigan checklist and LIDA scales are more suitable for the assessment of browser-based OHI (eg, the content layout of browser pages, use of browser navigation) [24,25], DISCERN was designed to evaluate the quality of online therapeutic information [6], and SMOG lacks support for Chinese-speaking users [26]. Therefore, in this study, we used the Suitability Assessment of Materials (SAM) scale to evaluate the health information released by HWPAs. The SAM scale comprises six dimensions, including content, literacy demand, graphics, layout and typography, learning stimulation/motivation, and cultural appropriateness, which have good reliability and validity and are widely used in evaluating the quality of OHI materials [27]. The SAM scale includes 22 factors for a total of 44 points (100%), with a higher score indicating better suitability. The results are rated on three levels: superior (70%-100%), adequate (40%-69%), and not suitable (0-39%).

## Rating Process

The evaluation was performed by three researchers. Rater 1 (WS) holds a master's degree in medical informatics and has 7 years of experience in medical information analysis and research. Rater 2 (FW) holds a master's degree in computer science and a doctorate degree in social medicine, with 8 years of experience in software development. Rater 3 (ZB) holds a doctorate degree and a clinician qualification, with 10 years of clinical experience. Pilot assessments were conducted using 6 HWPAs (3 certified and 3 uncertified). Before the pilot assessment, the three raters carefully studied the simplified Chinese description of the HONcode scale on the HON website and the English version of the SAM sheet to better understand the purpose and significance of each item. The assessment was performed in two steps. First, rater 1 and rater 2 evaluated the HWPAs' compliance with the principles of the HONcode. Second, rater 1 and rater 3 evaluated the suitability of the articles

released by the HWPAs with the SAM scale. To ensure consistency of the evaluation results, the evaluation was carried out with the subdimensions of the scale (ie, the Nth subdimension of all samples was completed, and then the Nth+1 dimension was evaluated). The assessment process was conducted in parallel; that is, two raters independently evaluated the same sample at the same time. For any controversial assessment results, the final results were determined through real-time negotiation.

### Statistical Analysis

Statistical analysis was conducted using the Statistical Package for Social Sciences version 17.0 (SPSS Inc, Chicago, IL, USA) and Excel version 2013 (Microsoft Inc, Washington DC, USA). All values are expressed as the mean (SD). Within-group comparisons of the SAM scores were performed using paired *t* tests. The critical value of significance was determined to be *P*=.05.

## Results

### Characteristics of the Health-Related WeChat Public Accounts

The characteristics of the HWPAs are shown in Table 1. Of the 93 HWPAs, 66 (71%) were officially verified by Tencent Corporation, while 27 (29%) were not. The WCI values ranged between 1479.35 and 877.62 and were divided into three sections: A (the first 20%), 1108.80-1479.35; B (middle 60%), 909.10-1100.52; and C (the last 20%), 877.62-907.74. The owners of certified HWPAs were mainly companies, whereas uncertified HWPAs were mainly owned by individuals. With respect to the type of information released, the certified and uncertified HWPAs were similarly dominated by comprehensive information, followed by traditional Chinese medicine (TCM).

**Table 1.** Summary of the descriptive and frequency statistics for the final sample of WeChat public accounts.

| Characteristic | Certified (N=66) | Uncertified (N=27) |
|---|---|---|
| **Subject classification, n (%)** | | |
| Company | 53 (80) | 3 (11) |
| NGO[a] | 12 (18) | 1 (4) |
| Individual | 1 (2) | 23 (85) |
| **WCI[b]** | | |
| Section A | 17 (26) | 3 (11) |
| Section B | 35 (53) | 19 (70) |
| Section C | 14 (21) | 5 (19) |
| **Type of content** | | |
| Traditional Chinese medicine | 5 (7) | 7 (26) |
| Rational diet | 2 (3) | 3 (11) |
| Sports and health | 2 (3) | 3 (11) |
| Mental health | 1 (2) | 0 (0) |
| Medical scientific research | 2 (3) | 0 (0) |
| Health of key population | 5 (8) | 0 (0) |
| Comprehensive information | 49 (74) | 14 (52) |

[a]NGO: nongovernmental organization.

[b]WCI: WeChat Communication Index.

### Health on the Net Foundation Code of Conduct Conformity

The HONcode compliance of the 93 HWPAs is shown in Multimedia Appendix 2. Although certified HWPAs were slightly more compliant than uncertified HWPAs, the compliance of HWPAs with the HONcode principles was generally not ideal, especially regarding the six principles of privacy, attribution, justifiability, transparency, financial disclosure, and advertising policy. For the remaining two principles, the compliance was also uneven. Although most HWPAs provided information on the content providers, they seldom verified the qualifications of these providers. In addition, most of the HWPAs did not state that the health-related information they provided was intended to support rather than replace medical decisions, and they did not clearly identify the user groups they were targeting.

### Suitability of Articles From WeChat Public Accounts

Among the six subdimensions of the SAM scale, layout and typography scored the highest (4.83/6 points), and graphics (5.167/10 points) and learning stimulation (3.68/6 points) scored the lowest. The remaining three subdimensions scored somewhere in between (7.22/10 points and 3.02/4 points), closer to the typical SAM score (34/44 points, 77%). There were 89 (42.4%) articles (mean 77.58, SD 5.496) that met the criteria

for “superior” suitability as established by the SAM scale, 118 (56.2%) articles (mean 61.06, SD 7.014) that met the criteria for “adequate” suitability, and 3 (1.4%) articles (mean 38.64, SD 2.143) that were evaluated as “not suitable”. The mean SAM score was 67.70 (SD 10.93), indicating “adequate” suitability.

The descriptive statistics of the SAM scale are shown in Table 2.

The results of the *t* test (Table 3) indicated that there were no significant differences in the SAM scores of the articles released by HWPAs with and without Tencent certification, except for the literacy requirements dimension.

**Table 2.** Descriptive statistics of the Suitability Assessment of Materials scale (N=210).

| Item | Not suitable | Adequate | Superior |
|---|---|---|---|
| **Content, n (%)** | | | |
| Purpose | 0 (0) | 23 (10.9) | 187 (89.1) |
| Content topics | 1 (0.5) | 96 (45.7) | 113 (53.8) |
| Scope | 3 (1.4) | 117 (55.7) | 90 (42.9) |
| Summary/review | 51 (24.3) | 100 (47.6) | 59 (28.1) |
| **Literacy demand, n (%)** | | | |
| Reading grade level | 60 (28.6) | 105 (50.0) | 45 (21.4) |
| Writing style | 28 (13.3) | 67 (31.9) | 115 (54.8) |
| Vocabulary | 29 (13.8) | 109 (51.9) | 72 (34.3) |
| Context | 1 (0.5) | 26 (12.4) | 183 (87.1) |
| Advanced organizers | 13 (6.2) | 14 (6.7) | 183 (87.1) |
| **Graphics, n (%)** | | | |
| Cover graphic | 42 (20.0) | 107 (51.0) | 61 (29.0) |
| Type of illustrations | 24 (11.4) | 95 (45.3) | 91 (43.3) |
| Relevance of illustrations | 41 (19.5) | 94 (44.8) | 75 (35.7) |
| List, tables, graphs, charts | 76 (36.2) | 73 (34.8) | 61 (29.0) |
| Captions | 103 (49.1) | 74 (35.2) | 33 (15.7) |
| **Layout and typography, n (%)** | | | |
| Layout | 0 (0) | 96 (45.7) | 114 (54.3) |
| Typography | 5 (2.4) | 89 (42.4) | 116 (55.2) |
| Subheadings | 20 (9.5) | 11 (5.2) | 179 (85.2) |
| **Learning stimulation/motivation, n (%)** | | | |
| Interaction | 139 (66.2) | 66 (31.4) | 5 (2.4) |
| Modeling of behaviors | 6 (2.9) | 48 (22.9) | 156 (74.3) |
| Motivation | 18 (8.6) | 47 (22.4) | 145 (69.1) |
| **Cultural appropriateness, n (%)** | | | |
| Cultural match | 1 (0.5) | 17 (8.1) | 192（91.4） |
| Cultural image and examples | 34 (16.2) | 119 (56.7) | 57 (27.1) |

**Table 3.** Evaluation scores of articles on the WeChat public accounts (mean, SD).

| SAM[a] item | Certified | Uncertified | *t* | *P* value |
|---|---|---|---|---|
| Content | 5.94 (1.244) | 5.70 (1.049) | –1.362 | .18 |
| Literacy demand | 7.06 (1.764) | 7.70 (1.612) | –2.418 | .02 |
| Graphics | 4.98 (2.683) | 5.72 (2.582) | 1.777 | .08 |
| Layout and typography | 4.77 (1.181) | 5.00 (1.019) | 1.263 | .21 |
| Learning stimulation, motivation | 3.70 (1.135) | 3.62 (1.023) | –0.466 | .64 |
| Cultural appropriateness | 2.99 (0.734) | 3.11 (0.670) | 1.155 | .25 |
| Total Score | 29.44(4.993) | 30.85(4.190) | 1.847 | .07 |

[a]SAM: Suitability Assessment of Materials.

# Discussion

## Principal Findings

The HWPAs had overall low HONcode conformity. Although the suitability of health information released by HWPAs was at a moderate level, there were still problems identified such as difficulty in tracing information sources, excessive implicit advertisements, and irregular usage of charts.

HONcode certification was significantly correlated with website quality as measured by DISCERN [28]. However, this survey found that the HONcode compliance of HWPAs was low, which is similar to the HONcode compliance of nonChinese health websites such as those on urinary diseases [29] and Ebola [30]. By analyzing the certificate information of HWPAs, we found that most of the certified HWPA owners were companies. These companies also have their own websites that are accessed through internet browsers in addition to the HWPAs, and the articles they pushed through HWPAs were also published on their websites. After further checking these Web page-based websites, we found that none of the websites included in this survey had HONcode certification. Although relevant studies on the HONcode compliance of nonChinese health websites have reported some cases of missing HONcode certification [28,31], the results of this survey are clearly more negative. The main approach to monitor the quality of online health information in China is to check the business scope and content of the network information platform according to relevant laws and regulations. This approach emphasizes the binding force of the legitimacy of the network information platform and the external monitoring of the online information platform. HONcode is not a strict regulation but is rather an ethical code for the release of online information that emphasizes the constraints on online information providers from the perspective of professional spirit, moral conscience, and internal monitoring of the network information platform [32]. In this survey, we found that the business entities of HWPAs paid more attention to compliance with laws and regulations but paid insufficient attention to the ethical standards for online information, such as privacy, attribution, justification, or transparency. This may be the main reason for the low HONcode conformity of HWPAs.

We also found a lack of appropriate mechanisms for monitoring the backgrounds of the HWPAs. Regardless of whether the HWPAs were operated by individuals or by organizations, there was no requirement to provide any medical qualification-related certification materials to apply for a WPA [33]. Although some organizations (such as hospitals, government agencies, and medical-related media) have business licenses, we can only infer whether they have medical-related qualifications [34] or internet content provider qualifications [35]; however, some trading companies, advertising companies, associations, and other organizations have also received Tencent's official WPA certification (see Multimedia Appendix 3). For personal applicants, as long as they submit identification information (eg, an identity card), a cell phone number, and a bank card tied to a WeChat ID, they can apply to establish their own HWPAs without the need to provide any proof of medical qualifications [33].

In addition, it is common to search for WPAs by name. However, the names of many WPAs were not matched to their WeChat IDs. For example, we searched for a WPA named 中医养生 (Traditional Chinese Medicine and Healthcare), and 11 results were returned. Although each WPA has a unique WeChat ID, it is difficult for consumers to remember the IDs or to distinguish among the WPAs using their IDs.

Regarding the overall evaluation, the HWPA health information suitability was determined to be at the "moderate" level, which is similar to the published OHI suitability evaluation results [36,37]. However, regarding literacy demand and cultural appropriateness, the HWPA scores were significantly better than those of some nonChinese OHI evaluation results [28]. This may be because, compared with China with a relatively stable cultural environment, countries in Europe and America have more immigrants and face more problems such as cultural assimilation [38], income disparity [39], and disease burden [40]. The resulting cultural and linguistic differences inevitably lead to differences in people's health-related behavior [41] and understanding of OHI [42]. On the one hand, this requires website owners to consider more acculturation factors when publishing health information. On the other hand, it creates higher requirements for users' cultural literacy [43].

In terms of scoring dimensions, most of the health-related articles published by HWPAs had friendly cover pictures and attractive titles that clearly described the purpose of the article, had a good layout and typography, and were culturally suitable. However, the use of charts was not standardized, and the lack of charts used as illustrations was a common problem. More

than half of the articles included pictures with little relevance to the content of the articles or even negative exaggerations and stereotypical cultural characteristics. In addition, in terms of the vocabulary used, readers would have little difficulty reading the articles, but articles related to TCM generally used professional terms. Although knowledge of TCM, as part of the national traditional culture, can be expected of most Chinese residents, there was still a large amount of content that would be difficult for readers to understand [44], indicating a higher level of literacy required to understand the articles.

In addition, compared with the traditional mode of health information dissemination through Web pages, WPAs require simple operations to share information, and the integrated payment functions in WeChat, Alipay, and other apps only require a few steps to complete commodity purchases and even provide short-term interest-free loan and installment repayment services. Therefore, article copying and implicit advertisements were common problems of the HWPAs. We randomly selected several articles, which were searched using the Baidu search engine, and found that some articles were posted on different websites; the same results were obtained for several articles marked as "original" on the HWPAs. Although most websites cited their sources, the sources indicated for the same articles on different websites were often inconsistent. This might explain why there was no significant difference in the results of the evaluation of the suitability of articles released by certified and uncertified HWPAs. Advertisements were usually embedded in the text in the form of pictures or articles introducing sales information for products in the form of an article summary. We assessed the accuracy of the health knowledge disseminated through several articles containing implicit advertisements and found no medicine-related errors. However, such implicit advertisements might still cause undesirable subjective feelings regarding user access, reduce consumers' trust in the content of the articles, or mislead consumers regarding healthy behaviors [45].

## Implications

New media has become an important resource for the public to seek health information. As representatives of Chinese health information communication platforms accessed through mobile phones and mobile terminals, HWPAs have a large consumer base in the Chinese-speaking world [46]. We suggest that the owners of HWPAs should follow the HONcode to guide and improve the construction of their HWPAs and strengthen the quality control of the OHI they publish. At the same time, as the manager of WeChat public account platforms, the Tencent Corporation should strengthen the qualification requirements for applying for an HWPA and strengthen the supervision of the health-related content released by WPAs to avoid the occurrence of another "Wei Zexi incident" [47] on WPA platforms. In addition, we suggest that related research institutions formulate targeted norms for the construction of mobile OHI platforms.

## Limitations

Several limitations of this study are apparent. First, there are many evaluation indices of WPAs, but there is a lack of horizontal comparison of these indices. In this study, we chose the WCI proposed by Qingbo Bigdata as the ranking basis for the influence of WPAs, which may have resulted in selection bias. Second, as a set of principles that health websites should follow, the HONcode is important in guiding the construction of HWPAs. However, a few indicators were not well targeted to mobile platforms, which may have reduced the validity of the assessment. Third, the Chinese nation is a multiethnic group, and some ethnic minorities have their own spoken and written languages. However, due to the WCI ranking system, the content on the HWPAs in this survey was all in simplified Chinese. Finally, all evaluations were influenced by the researchers who conducted them, and the results of their evaluations may differ from consumers' feelings. In view of the above limitations, the conclusions of this study are preliminary and should be carefully interpreted.

## Conclusions

We found that HWPAs had low compliance with the HONcode. Although the suitability of the articles released by HWPAs was at a moderate level, there were still problems identified, such as difficulty in tracing the sources of information, excessive implicit advertisements, and the irregular usage of charts. Moreover, low approval requirements for applications to obtain an HWPA are not conducive to improving the service quality of HWPAs.

## Acknowledgments

This study was supported by funding from the National Planning Office of Philosophy and Social Science of the People's Republic of China (grant no. 17BGL262). The authors wish to thank American Journal Expert (AJE) for English language editing.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
WCI ranking data for WPAs (March 1-31, 2019).
[XLSX File (Microsoft Excel File), 38 KB - mhealth_v8i5e14826_app1.xlsx ]

Multimedia Appendix 2
Descriptive statistics of HONcode conformity (n, %).

[DOCX File , 16 KB - mhealth_v8i5e14826_app2.docx ]

Multimedia Appendix 3
Raw data of HWPAs rating scale.
[XLS File (Microsoft Excel File), 213 KB - mhealth_v8i5e14826_app3.xls ]

## References


1. Zai D, Anna D, Jeannette I, Trace K. Tweeting About Testing: Do Low-Income, Parenting Adolescents and Young Adults Use New Media Technologies to Communicate About Sexual Health? Perspect Sex Reprod Health 2012 Sep;44(3):176-183. [doi: 10.1363/4417612] [Medline: 22958662]
2. Shaw RJ, Johnson CM. Health Information Seeking and Social Media Use on the Internet among People with Diabetes. Online J Public Health Inform 2011 Jun;3(1):ojphi.v3i1.3561. [doi: 10.5210/ojphi.v3i1.3561] [Medline: 23569602]
3. Hou XR, Chen JY, Zhao WL. Quality analysis of public medical and health information on the internet. Chinese Journal of Health Informatics and Management 2014;11(1):38-42. [doi: 10.3969/j.issn.1672-5166.2014.01.07]
4. Qian MH, Xu ZX, Wang S. Research on information service quality of online health platform based on user participation. Journal of the China Society for Scientific and Technical Information 2019;38(2):132-142.
5. Deng SL, Zhao HP. Research on the standard framework of the quality and the content evaluation of online health information from user's perspective. Library Inform Serv 2017(21):30-39. [doi: 10.13266/j.issn.0252-3116.2017.21.004]
6. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
7. Team HON. Health On the Net Foundation. HONcode certification (Introduction) URL: https://www.hon.ch/HONcode/ [accessed 2019-04-20]
8. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997 Apr 16;277(15):1244-1245. [Medline: 9103351]
9. Tomlin ABD. The Minervation validation instrument for healthcare websites (LIDA instrument v1.2). URL: http://www.minervation.com/wp-content/uploads/2011/04/Minervation-LIDA-instrument-v1-2.pdf [accessed 2020-03-29]
10. Deng SL, Zhao HP. Quality evaluation of foreign network health information: A review of indicators, tools and results. Inform Document Serv 2017(1):67-74.
11. Yu WW, Wang K, Cao JD. Comparison between Chinese and American internet health information service websites. 2015 Jun 18 Presented at: The 21st National Medical Informatics Academic Conference of Chinese Medical Association; 2015; Zhengzhou.
12. Taipingyang Network. Shenzhen Tencent Computer System Co Ltd. 2019 Jan 9. WeChat releases 2018 data report URL: https://xw.qq.com/amphtml/20190109A0Q6MJ00 [accessed 2019-01-19]
13. Zhang L. Health education research based on WeChat public platform. J Nurs Rehab 2018(8):84-86. [doi: 10.3969/j.issn.1671-9875.2018.08.028]
14. Wang CJ. Eliminate false, keep true，the function of health-related WeChat public account in science population. New Media Research 2018;4(19):128-129. [doi: 10.16604/j.cnki.issn2096-0360.2018.19.055]
15. Jin Y, Zhang QY. Research on factors affecting the attention degree on health public WeChat account: based on root theory. Library Theor Pract 2018(4):48-52. [doi: 10.14064/j.cnki.issn1005-8214.2018.04.011]
16. Chi H. The spread and governance of health rumours disseminated in Wechat Public Account. Research on Digital Media 2018;35(5):27-30.
17. Li W, Han LQ, Guo YJ, Sun J. Using WeChat official accounts to improve malaria health literacy among Chinese expatriates in Niger: an intervention study. Malar J 2016 Nov 24;15(1):567 [FREE Full text] [doi: 10.1186/s12936-016-1621-y] [Medline: 27881122]
18. Guo Y, Xu Z, Qiao J, Hong YA, Zhang H, Zeng C, et al. Development and Feasibility Testing of an mHealth (Text Message and WeChat) Intervention to Improve the Medication Adherence and Quality of Life of People Living with HIV in China: Pilot Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 Sep 04;6(9):e10274 [FREE Full text] [doi: 10.2196/10274] [Medline: 30181109]
19. Cao Y, Lin S, Zhu D, Xu F, Chen Z, Shen H, et al. WeChat Public Account Use Improves Clinical Control of Cough-Variant Asthma: A Randomized Controlled Trial. Med Sci Monit 2018 Mar 14;24:1524-1532 [FREE Full text] [doi: 10.12659/msm.907284] [Medline: 29536984]
20. Beijing Qingbo Bigdata Technology Co Ltd. Qingbo Bigdata Company Profile URL: http://home.gsdata.cn/about-us.html [accessed 2019-10-28]
21. Qingbo Bigdata Technology Co Ltd. Measurement description of WCI Beijing URL: http://www.gsdata.cn/site/usage-1 [accessed 2019-10-28]

22. Boyer C, Dolamic L. Automated Detection of HONcode Website Conformity Compared to Manual Detection: An Evaluation. J Med Internet Res 2015 Jun 02;17(6):e135 [FREE Full text] [doi: 10.2196/jmir.3831] [Medline: 26036669]
23. Team HON. Health On the Net Foundation. mHONcode: The new certification of health mobile applications URL: https://www.hon.ch/en/certification/app-certification-en.html [accessed 2019-05-10]
24. Anderson PF, Allee N, Grove S, Hill S. Ann Arbor: University of Michigan. 1999. Web site evaluation checklist URL: http://www-personal.umich.edu/~pfa/pro/courses/WebEvalNew.pdf [accessed 2019-03-22]
25. Tanabe K, Fujiwara K, Ogura H, Yasuda H, Goto N, Ohtsu F. Quality of Web Information About Palliative Care on Websites from the United States and Japan: Comparative Evaluation Study. Interact J Med Res 2018 Apr 03;7(1):e7 [FREE Full text] [doi: 10.2196/ijmr.9574] [Medline: 29615388]
26. Luk A, Aslani P. Tools used to evaluate written medicine and health information: document and user perspectives. Health Educ Behav 2011 Aug;38(4):389-403. [doi: 10.1177/1090198110379576] [Medline: 21490309]
27. Robins S, Barr HJ, Idelson R, Lambert S, Zelkowitz P. Online Health Information Regarding Male Infertility: An Evaluation of Readability, Suitability, and Quality. Interact J Med Res 2016 Oct 21;5(4):e25 [FREE Full text] [doi: 10.2196/ijmr.6440] [Medline: 27769954]
28. Murray KE, Murray TE, O'Rourke AC, Low C, Veale DJ. Readability and Quality of Online Information on Osteoarthritis: An Objective Analysis With Historic Comparison. Interact J Med Res 2019 Sep 16;8(3):e12855 [FREE Full text] [doi: 10.2196/12855] [Medline: 31538953]
29. Wong L, Yan H, Margel D, Fleshner NE. Urologists in cyberspace: A review of the quality of health information from American urologists' websites using three validated tools. Can Urol Assoc J 2013 Apr;7(3-4):100-107 [FREE Full text] [doi: 10.5489/cuaj.501] [Medline: 23671523]
30. Hamzehei R, Ansari M, Rahmatizadeh S, Valizadeh-Haghi S. Websites as a tool for public health education: determining the trustworthiness of health websites on Ebola disease. Online J Public Health Inform 2018;10(3):e221 [FREE Full text] [doi: 10.5210/ojphi.v10i3.9544] [Medline: 30680054]
31. Valizadeh-Haghi S, Rahmatizadeh S. Evaluation of the Quality and Accessibility of Available Websites on Kidney Transplantation. Urol J 2018 Sep 26;15(5):261-265 [FREE Full text] [doi: 10.22037/uj.v0i0.4252] [Medline: 30058062]
32. Zhang HH, Ma JD, Di JP. Assessment on Information Quality of Health Websites. J Med Informatics 2014 Mar 20;35(3):2-5. [doi: 10.3969/j.issn.1673-6036.2014.03.001]
33. Tencent Customer Service Centre. Shenzhen Tencent Computer System Co Ltd. What information dose the registration of WeChat Public Account need to prepare? URL: http://kf.qq.com/touch/faq/150722mYN3IB150722yEreeQ.html?platform=15 [accessed 2019-04-29]
34. State Administration for Market Regulation, PRC. National Enterprise Credit Information Publicity System URL: http://www.gsxt.gov.cn/index.html [accessed 2019-05-11]
35. Ministry of Industry and Information Technology, PRC. ICP/IP address/Domain name information filing management system URL: https://tinyurl.com/y84cxvgm [accessed 2019-05-11]
36. Tran BNN, Singh M, Lee BT, Rudd R, Singhal D. Readability, complexity, and suitability analysis of online lymphedema resources. J Surg Res 2017 Jun 01;213:251-260. [doi: 10.1016/j.jss.2017.02.056] [Medline: 28601323]
37. Monton O, Lambert S, Belzile E, Mohr-Elzeki D. An evaluation of the suitability, readability, quality, and usefulness of online resources for family caregivers of patients with cancer. Patient Educ Couns 2019 Oct;102(10):1892-1897. [doi: 10.1016/j.pec.2019.05.010] [Medline: 31151780]
38. Asgari H, Zaman N, Jin X. Understanding Immigrants' Mode Choice behavior in Florida: Analysis of Neighborhood Effects and Cultural Assimilation. Transportation Research Procedia 2017;25:3079-3095. [doi: 10.1016/j.trpro.2017.05.319]
39. Bárcena-Martín E, Pérez-Moreno S. Immigrant–native gap in poverty: a cross-national European perspective. Rev Econ Household 2016 Jan 9;15(4):1105-1136. [doi: 10.1007/s11150-015-9321-x]
40. Lenzi J, Avaldi VM, Rucci P, Pieri G, Fantini MP. Burden of multimorbidity in relation to age, gender and immigrant status: a cross-sectional study based on administrative data. BMJ Open 2016 Dec 21;6(12):e012812. [doi: 10.1136/bmjopen-2016-012812] [Medline: 28003289]
41. Shapiro E. Places of Habits and Hearts: Church Attendance and Latino Immigrant Health Behaviors in the United States. J Racial Ethn Health Disparities 2018 Dec;5(6):1328-1336. [doi: 10.1007/s40615-018-0481-2] [Medline: 29549652]
42. Neumark Y, Flum L, Lopez-Quintero C, Shtarkshall R. Quality of online health information about oral contraceptives from Hebrew-language websites. Isr J Health Policy Res 2012;1(1):38 [FREE Full text] [doi: 10.1186/2045-4015-1-38] [Medline: 23006798]
43. Shen LN, Wang SM, Chen WQ, Fu Q, Evans R, Lan FQ, et al. Understanding the Function Constitution and Influence Factors on Communication for the WeChat Official Account of Top Tertiary Hospitals in China: Cross-Sectional Study. J Med Internet Res 2019 Dec 09;21(12):e13025 [FREE Full text] [doi: 10.2196/13025] [Medline: 31815674]
44. Tian WW. Investigation study on the cognition, attitudebehaviour of community residents on nutrition of traditional Chinese medicine. M. A. Thesis. Beijing: Beijing University of Chinese Medicine 2014.
45. Wang FZ, Luo AJ, Luo D, Hu DH, Xie WZ, Chen ZH, et al. The association between health information attitudes and skills in patients with chronic disease in China. Glob Health Promot 2016 Dec 09;23(4):59-69. [doi: 10.1177/1757975915577922] [Medline: 26041814]

46. Zhang X, Wen D, Liang J, Lei J. How the public uses social media wechat to obtain health information in china: a survey study. BMC Med Inform Decis Mak 2017 Jul 05;17(Suppl 2):66 [FREE Full text] [doi: 10.1186/s12911-017-0470-0] [Medline: 28699549]
47. People's Daily Online. 2016 May 5. Op-ed: Wei Zexi's story highlights the importance of proactive regulation URL: http://en.people.cn/n3/2016/0505/c98649-9053533.html [accessed 2020-04-02]

## Abbreviations

**HONcode:** Health on the Net Foundation Code of Conduct
**HWPA:** health-related WeChat public account
**OHI:** online health information
**SAM:** Suitability Assessment of Material
**SMOG:** Simple Measure of Gobbledygook
**TCM:** traditional Chinese medicine
**WCI:** WeChat Communication Index
**WPA:** WeChat public account

*Edited by G Eysenbach; submitted 27.05.19; peer-reviewed by M Rethlefsen, J Liang; comments to author 03.10.19; revised version received 19.11.19; accepted 12.03.20; published 08.05.20.*

*Please cite as:*
*Wang F, Wang Z, Sun W, Yang X, Bian Z, Shen L, Pan W, Liu P, Chen X, Fu L, Zhang F, Luo D*
*Evaluating the Quality of Health-Related WeChat Public Accounts: Cross-Sectional Study*
*JMIR Mhealth Uhealth 2020;8(5):e14826*
*URL: https://mhealth.jmir.org/2020/5/e14826*
*doi:10.2196/14826*
*PMID:32383684*

Original Paper

# A Questionnaire for Assessing User Satisfaction With Mobile Health Apps: Development Using Rasch Measurement Theory

Jeanette Melin[1], PhD; Stephanie Erika Bonn[2], PhD; Leslie Pendrill[1], PhD; Ylva Trolle Lagerros[2], MPH, MD, PhD

[1]Research Institutes of Sweden AB, Göteborg, Sweden

[2]Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

**Corresponding Author:**
Ylva Trolle Lagerros, MPH, MD, PhD
Clinical Epidemiology Division
Department of Medicine Solna
Karolinska Institutet
Eugeniahemmet T2
Stockholm, 171 76
Sweden
Phone: 46 46 851779183
Fax: 46 46 851779304
Email: ylva.trolle@ki.se

## Abstract

**Background:** Mobile health (mHealth) apps offer great opportunities to deliver large-scale, cost-efficient digital solutions for implementing lifestyle changes. Furthermore, many mHealth apps act as medical devices. Yet, there is little research on how to assess user satisfaction with an mHealth solution.

**Objective:** This study presents the development of the mHealth Satisfaction Questionnaire and evaluates its measurement properties.

**Methods:** Respondents who took part in the Health Integrator Study and were randomized to use the Health Integrator smartphone app for lifestyle changes (n=112), with and without additional telephone coaching, rated their satisfaction with the app using the new 14-item mHealth Satisfaction Questionnaire. The ratings were given on a 5-point Likert scale and measurement properties were evaluated using Rasch measurement theory (RMT).

**Results:** Optimal scoring was reached when response options 2, 3, and 4 were collapsed, giving three response categories. After omitting two items that did not fit into the scale, fit residuals were within, or close to, the recommended range of ±2.5. There was no differential item functioning between intervention group, age group, or sex. The Person Separation Index was 0.79, indicating that the scale's ability to discriminate correctly between person leniency was acceptable for group comparisons but not for individual evaluations. The scale did not meet the criterion of unidimensionality; 16.1% (18/112) of the respondents were outside the desired range of −1.96 to 1.96. In addition, several items showed local dependency and three underlying dimensions emerged: negative experiences, positive experiences, and lifestyle consequences of using the mHealth solution.

**Conclusions:** In times where mHealth apps and digital solutions are given more attention, the mHealth Satisfaction Questionnaire provides a new possibility to measure user satisfaction to ensure usability and improve development of new apps. Our study is one of only a few cases where RMT has been used to evaluate the usability of such an instrument. There is, though, a need for further development of the mHealth Satisfaction Questionnaire, including the addition of more items and consideration of further response options. The mHealth Satisfaction Questionnaire should also be evaluated in a larger sample and with other mHealth apps and in other contexts.

**Trial Registration:** ClinicalTrials.gov NCT03579342; http://clinicaltrials.gov/ct2/show/NCT03579342.

***(JMIR Mhealth Uhealth 2020;8(5):e15909)*** doi:10.2196/15909

**KEYWORDS**

cell phone; healthy lifestyle; methods; mobile applications; psychometrics; smartphone; telemedicine; mobile phone

## Introduction

### Background

Electronic health (eHealth), defined by the World Health Organization as the use of information and communication technologies for health, also encompasses mobile health (mHealth), defined as a medical or public health practice that is supported by mobile devices [1]. mHealth has had a rapid evolution and adoption and today, smartphone apps have the potential to make the treatment and prevention of diseases cost-efficient and widely accessible. There is a vast number of apps for tracking different types of health data such as physical activity, diet, sleep, stress, and more. Apps also commonly act as medical devices or accessories to medical devices to, for example, apps diagnose heart rhythm abnormalities or function together with a glucose meter used by an insulin-dependent patient with diabetes. The mHealth literature has grown rapidly over the last couple of years and has so far primarily focused on the effectiveness and efficiency of the usability of mHealth solutions.

While effectiveness refers to the completeness of specified goals, such as improved health status, which is often measured in terms of medical examinations and self-reported health, efficiency on the other hand, relates to the resources used for accomplishment, such as cost for personnel or digital solutions, rather than to its implications. However, from a full usability perspective, effectiveness and efficiency are important, but given the often high attrition rates in mHealth studies [2], user satisfaction may be key for retention.

To enable more comprehensive evaluations of mHealth solutions, there is a need to develop a questionnaire to assess user satisfaction. To the best of our knowledge, there is no generic commonly available tool, for example, a questionnaire, for capturing user satisfaction with mHealth solutions using an app. However, King et al [3] constructed a user satisfaction survey following their 8-week feasibility testing of different physical activity apps. It consisted of 22 items asking the respondents to rate usability on a 6-point Likert-type scale. The survey was subsequently adapted by Mummah et al [4] into a 21-item questionnaire in which participants were asked to rate their level of agreement or disagreement with different statements on a 5-point Likert-type scale. This questionnaire was used to measure user satisfaction of Vegathon, an app aiming to increase vegetable intake among adults with obesity. Statements in the questionnaires by King et al [3] and Mummah et al [4], as well as in the adapted version used here, include areas of usability such as time consumption, motivation, understanding, and willingness to recommend it.

### Objectives

In this work, we have developed a generic 14-item version called the mHealth Satisfaction Questionnaire. We used the new questionnaire in a large randomized controlled mHealth trial, the Health Integrator Study [5]. Here, we report how we used psychometric Rasch measurement theory (RMT) to assess the ability of our new mHealth Satisfaction Questionnaire to measure user satisfaction with an mHealth solution. On the basis of the results, we also suggest improvements for a future mHealth Satisfaction Questionnaire.

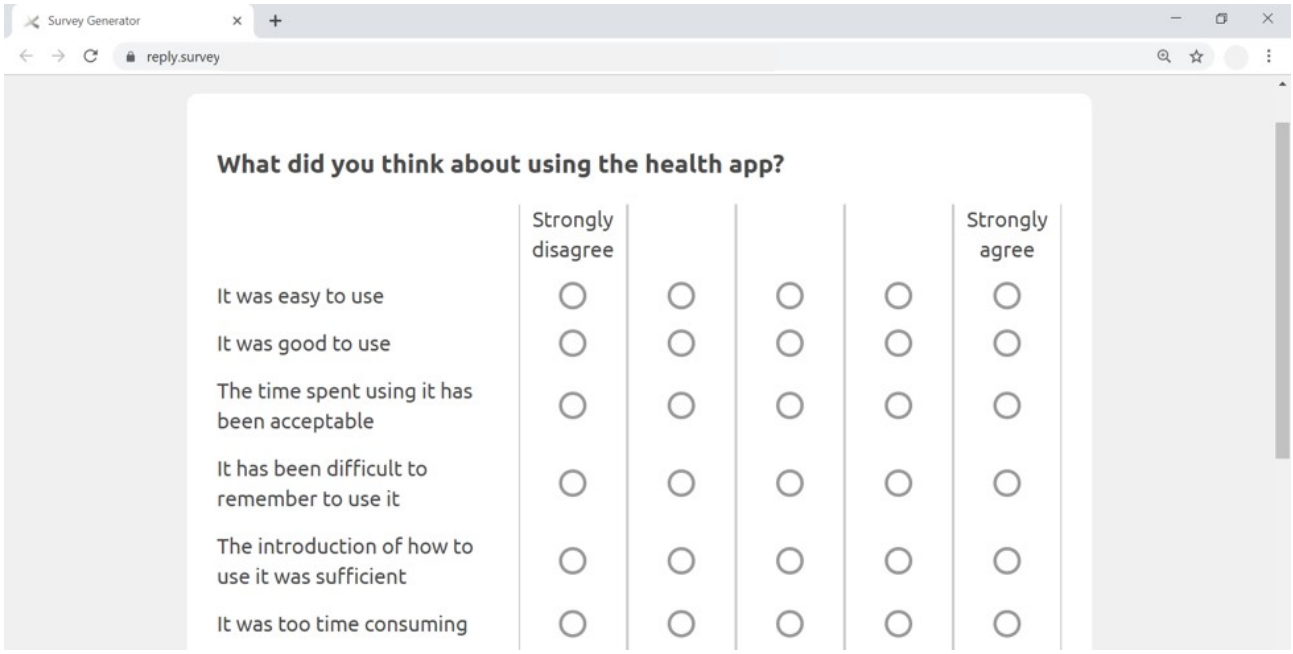
## Methods

### Data Collection

The study population comprised the respondents taking part in the Health Integrator Study's intervention groups (n=138). Briefly, the 3-month interventions included a personalized mHealth intervention based on the participant's personal health profile and were tailored to the need of each specific participant, with or without telephone sessions with a health coach. The intervention has been described in detail elsewhere [5]. The study was approved by the Regional Ethical Review Board in Stockholm, Sweden (2018/411-31 and 2018/1038-32).

At the 3-month follow-up of the active mHealth intervention, the participants were administered the Web-based mHealth Satisfaction Questionnaire, evaluating how the Health Integrator app was perceived—both in terms of usage and improved health. An example of the Web-based version of the questionnaire is shown in Figure 1. A reminder was sent after 2 weeks to all nonresponders.

**Figure 1.** Screenshot of the Web-based mHealth Satisfaction Questionnaire.




In total, 112 respondents completed the new mHealth Satisfaction Questionnaire after the intervention, corresponding to a response rate of 81%. Of these 112 respondents, 60 (53.5%) received telephone support with a health coach during the active intervention, while 52 (46.4%) did not receive this support. The study population comprised a slightly greater number of men, 56.3% (63/112), than women, 43.8% (49/112). The mean age of participants was 47.8 years (median 48.5 years, range 26-73 years).

## Measurement

The mHealth Satisfaction Questionnaire is an adaption of the user satisfaction survey used in the Vegathon study by Mummah et al [4], which in turn is an adaption of the user satisfaction survey by King et al [3]. In adjusting the questionnaire to a short and more generic 14-item version, we omitted some of the more vegetable-specific questions used in the Vegathon study such as *Vegathon has given me the confidence that I could become a better vegetable eater*. Although Mummah et al [4] specifically targeted the vegetable intake app in each statement, our questionnaire is divided into two sections with the overarching questions: *What did you think about using the health app?* and *How did you experience the health app?* The sections of our questionnaire include general statements about, for example, the usability of and willingness to recommend the app.

The mHealth Satisfaction Questionnaire consists of 14 items where the respondent is asked to rate to what extent he or she agrees on each item on a 5-point Likert-scale (see Multimedia Appendix 1). Higher rating corresponds to higher agreement (ie, 1=strongly disagree, 5=strongly agree); 10 items are positively stated, while four items are negatively stated. The negatively stated items were reversed in the analyses, and consequently, higher values correspond to higher leniency.

Measurements made with any questionnaire need to produce results which are invariant and reliable in much the same way as is required of physical measurements, for instance of mass or length [6]. To ensure equitability and fairness, irrespective of to whom the questionnaire is administered, estimates of person and item characteristics (leniency and quality in this case) deduced from questionnaire responses need to be comparable, as far as possible, with corresponding estimates made on other occasions by other independent measurements. Second, any measurement result will have limited quality as there is neither time nor resources to perform perfect and complete measurements. The risks of incorrect decisions (about for instance care) associated with this uncertainty can be assessed if measurement reliability is openly declared in terms of how much uncertainty there is in the actual measurement results at hand.

Invariant and reliable quality-assured measurement results based on questionnaires require principally two actions when analyzing responses: (1) raw data from questionnaires (here: respondents rated degree of agreement to each statement) are always ordinal [7], and need to be transformed on to a common measurement interval scale on which distances have quantitative meaning [8,9], and (2) raw questionnaire response data are a mix of person leniency and item quality, which need to be estimated separately.

RMT is a means of performing both of these actions: measurement data on interval (in contrast to ordinal) scales can be reliably analyzed with all the regular statistical tools and metrics, and separate estimates of person leniency and item quality enable metrological references for comparability, in much the same way as mass standards can only be established with separately calibrated weighing machines [10].

RMT was developed by the Danish mathematician Georg Rasch in the mid-20th century with the intention to enable invariant and individual comparisons, based on the same underlying principles as physical measurements [8]. In metrological terms, RMT can be understood as modeling a measurement system in which each item of a questionnaire represents a characteristic of an object to be measured (here: the quality demand value δ),

and each person responding to the questionnaire acts as a measurement instrument (with a certain respondent leniency θ) when providing a system response to the questionnaire (here: respondents' rated degree of agreement to each statement) [9].

According to RMT, data are evaluated against a mathematical model for guiding the construction of stable linear measures from raw data [8]. In the simplest, dichotomous case, transformation of questionnaire responses is made with a logistic regression function where $P_{success}$ is the probability of making a *correct* binary response:

Rasch modeled the log-odds of a *yes or pass* response to a *no or fail* response as a simple linear difference between person leniency (θ) and item quality (δ). This dichotomous model can be extended to a polytomous model (ie, Likert scales with multiple ratings) [11], which has been used in this study.

### Statistical Analysis

The measurement properties of the new mHealth Satisfaction Questionnaire were evaluated according to RMT, which estimated each participant's level of leniency (how easily satisfied they were) and each item's level of quality (the ability of each aspect of the app to make the user satisfied) by making a logistic regression to the complete set of responses of the whole cohort to all items of the questionnaire. This was done with the software Rasch Unidimensional Measurement Model (RUMM) 2030. To ensure that this logistic regression satisfied fundamental measurement properties, the analysis focused on the requirements *response category functioning*, *targeting and reliability,* and *model fit* [12,13].

#### *Response Category Functioning*

To evaluate the monotonicity of item response categories, the threshold orders were evaluated. Ratings on each item should be consistent with the metric estimate of the underlying construct, that is, ordered from low to high degree of agreement. This was completed as a first step, and where needed, categories were collapsed when disordered thresholds occurred [13].

#### *Targeting and Reliability*

Person locations should ideally mirror the item locations. Comparing the mean person location with the mean item location (ie, 0 logits) gives an indication to whether a person is off centered from the items [13]. Moreover, to evaluate the ability to successfully furnish separate estimates of each respondent's level of leniency, the Person Separation Index (PSI) was used to estimate reliability. Reliability, that is, the scale's ability to discriminate correctly between person leniency, was interpreted as follows: zero (0) indicates total uncertainty and one (1) implies no uncertainty, a result >0.70 is required for group assessments and >0.85 for individual high-stake evaluations items [13,14]. A person separation reliability of 0.8 indicates that measurement uncertainty is not more than one half of the total standard deviation observed [15].

#### *Model Fit*

Several fit statistics were evaluated including; fit residuals, $\chi^2$, item characteristic curve (ICC), differential item functioning (DIF), local dependency, and unidimensionality. We used the following guidelines:

1. A residual is the difference between a person's observed score on each item of the mHealth Satisfaction Questionnaire and the expected value derived from the RMT analysis. The mean residual is recommended to be close to zero (0) and standard deviations (SDs) close to one (1). At the same time, the individual item fit residuals should be within the range of −2.50 to +2.50 [12].
2. $\chi^2$ tests, which evaluate the difference between the observed and expected item responses, should ideally not be statistically significant (after Bonferroni correction) [11,12].
3. ICCs are graphical indicators of fit. They can be used to complement the interpretation of the fit residuals and $\chi^2$ probabilities. For ICC graphs, the dots of the class intervals should follow the ICC to support good fit [12].
4. DIF analyses are used to evaluate to which extent item responses are influenced by external factors, that is, item function should be similar across different groups and should ideally be nonsignificant (after Bonferroni correction) [16]. Both uniform and nonuniform DIF were tested for intervention group, age group, and gender. Age groups were created according to the following: *younger* corresponded to ≤39 years; *middle age* corresponded to 40 to 54 years; and *older* corresponded to ≥55 years.
5. Local dependency was evaluated according to a relative cut off of 0.2 above the average correlation [17,18]. To deal with local dependency, sets of items were grouped into new polytomous items, that is, *super items* with scores ranging from zero (0) to the maximum of the sum of the scores of the included items [19].
6. The Smith method for testing unidimensionality was applied [20]. This means that the first residual factor obtained in a principal component analysis is used to define two subsets of items by dividing them into positively and negatively correlated items. Thereafter, person estimates for each subset were compared using an independent *t* test. To support unidimensionality, the percentage of respondents outside the range −1.96 to 1.96 should not exceed 5%.

## Results

### 14 Item Version

We found disordered thresholds for all except one item in the questionnaire (*It was good to use*). This was, however, resolved by collapsing the response options. The optimal rescoring occurred when response categories 2, 3, and 4 were collapsed into one category. This was also done for the item with no disordered thresholds, as category probability curves showed close to disordered thresholds. Consequently, a 3-step scale was used for the remaining analyses, which was similar for all 14 items.

Comparisons of mean person location (0.60 logits; SD 1.13) and the mean item location (fixed to 0 logits, SD 0.71) indicated

that most persons' leniency levels were off centered with respect to the item quality locations. The PSI was 0.79, that is, the scale's ability to discriminate correctly between person leniency was acceptable for group comparisons but not for individual evaluations. In analysis of item hierarchy, the items were ordered in a logical line from the easiest demands for product quality (eg, *It was too time consuming* or *It was easy to use*) to the more demanding qualities of the product (eg, *It has helped me to understand the benefits of improving my lifestyle habits*).

The mean of the fit residuals was −0.45 (SD 1.4), indicating room for improvement in terms of fit to the RMT model. Nevertheless, the item fit statistics were satisfactory for all except two items, *It has been difficult to remember to use it* and *It interrupted me in my daily activities* (fit residuals 5.25 and 3.98, respectively; Table 1). None of the items showed statistically significant $\chi^2$, although, by studying the ICC, deviating dots were present for the two items with high fit residuals (Figures 2 and 3).

**Table 1.** Summary item statistics of the analyses for the version with 14 items.

| Items | Location | 2 SE | Fit residuals | Chi-square (*df*=2) | *P* value |
|---|---|---|---|---|---|
| It was a disturbance | −1.12 | 0.36 | −0.03 | 1.3 | .52 |
| It interrupted me in my daily activities | −1.00 | 0.35 | *3.98* [a] | 9.5 | .01 |
| It was too time consuming | −0.94 | 0.35 | −0.17 | 0.5 | .79 |
| The introduction of how to use it was sufficient | −0.45 | 0.37 | 1.14 | 3.2 | .21 |
| It was boring to use | −0.36 | 0.30 | 0.61 | 1.4 | .50 |
| It was easy to use | −0.29 | 0.36 | −0.83 | 2.2 | .33 |
| It was good to use | −0.01 | 0.43 | −1.16 | 4.6 | .10 |
| The time spent using it has been acceptable | 0.12 | 0.40 | −0.70 | 0.8 | .68 |
| I can recommend it to others | 0.40 | 0.36 | −0.87 | 2.7 | .26 |
| It has been difficult to remember to use it | 0.48 | 0.27 | *5.23* | 9.6 | .01 |
| It has motivated me to change my lifestyle habits | 0.72 | 0.38 | −1.91 | 5.8 | .06 |
| It has helped me to understand the benefits of improving my lifestyle habits | 0.74 | 0.39 | −1.99 | 5.3 | .07 |
| It has helped me to understand how I need to change my lifestyle habits | 0.78 | 0.38 | −2.20 | 4.1 | .13 |
| It has helped me set personal goals for my lifestyle habits in a way that I could not have done on my own | 0.94 | 0.37 | −1.55 | 3.1 | .21 |

[a]Fit residuals in *italic* indicate misfit +2.5.

**Figure 2.** Item characteristic curve (ICC) showing a line with the expected response (predicted from the model) and the dots corresponding to the observed response. The illustration shows how the dots deviated from the ICC for the item *It has been difficult to remember to use it.*

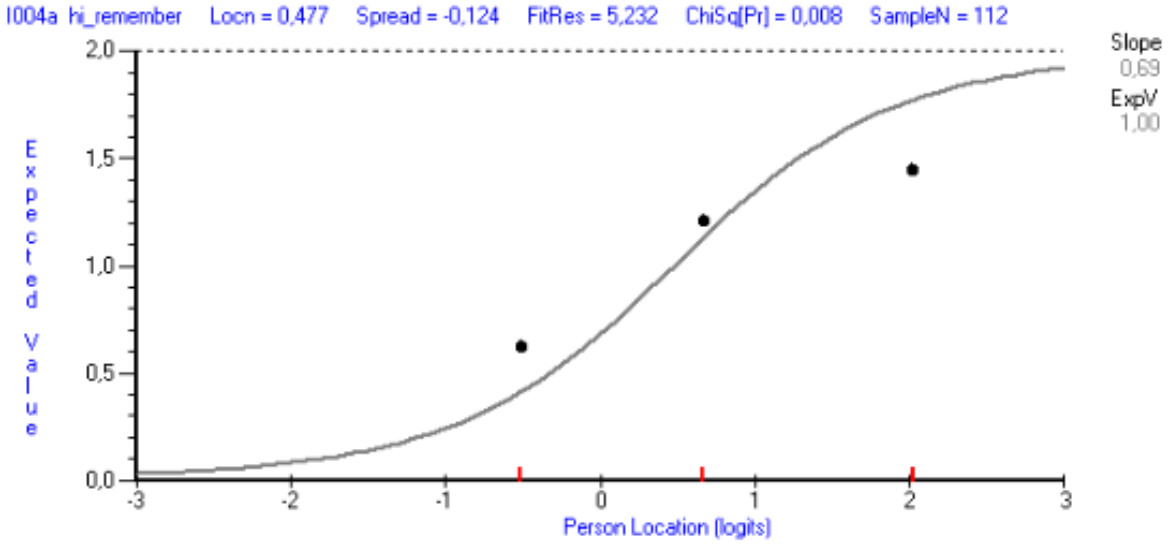
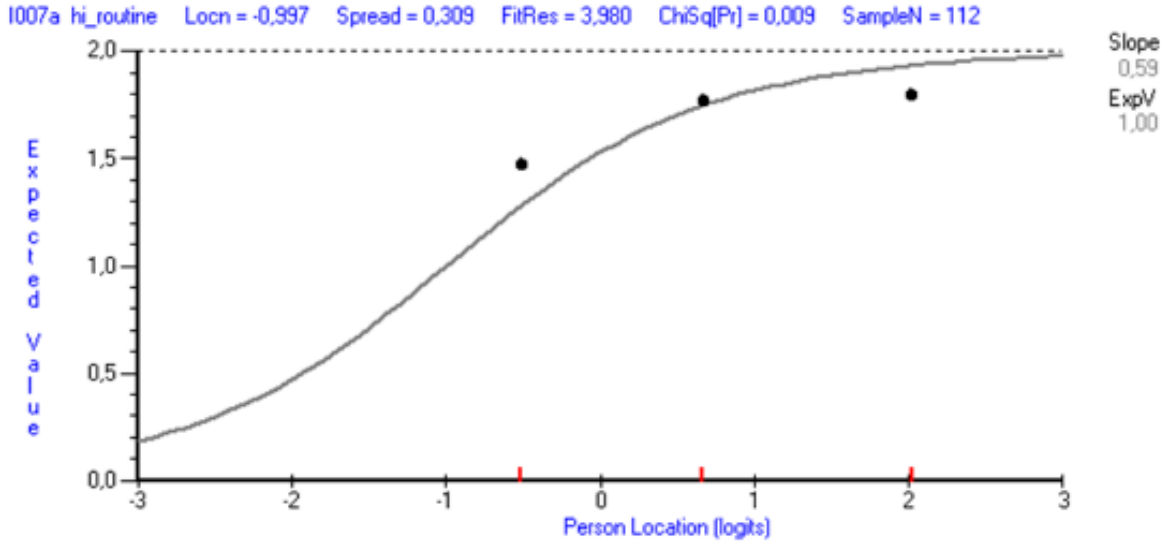


**Figure 3.** Item characteristic curve (ICC) showing a line with the expected response (predicted from the model) and the dots corresponding to the observed response. The illustration shows how the dots deviated from the ICC for the item *It interrupted me in my daily activities.*




There was no significant DIF present, neither uniform nor nonuniform, for any of the person factors (intervention group, age group, or gender). In total, 16 of 91 residual correlations failed to meet the relative cut off (0.14). A clear pattern of three distinct clusters was apparent, reflecting negative experiences, positive experiences, and lifestyle consequences of using the mHealth solution, respectively (Table 2). In addition, local dependency was also shown between the two items with unsatisfactory fit residuals. A *t* test revealed that 17.9% (20/112) of the respondents were outside the desired range of −1.96 to 1.96.

**Table 2.** Summary item statistics of the analyses for the version with 12 items.

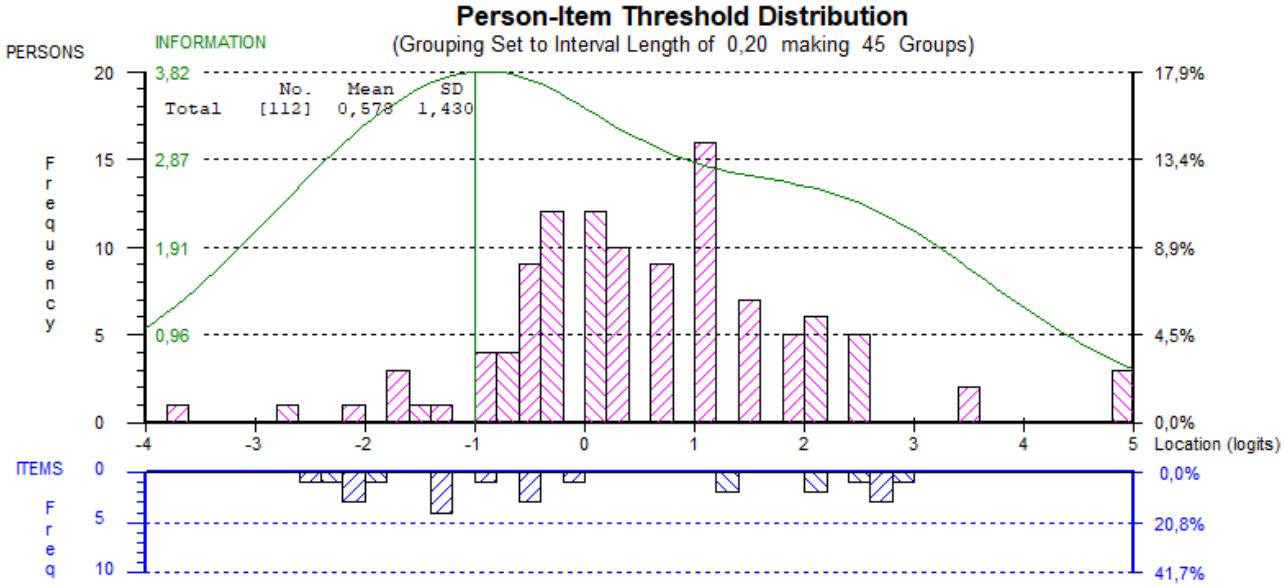
| Items | Location | 2 SE | Fit residuals | Chi-square (*df*=2) | *P* value | Testlet |
|---|---|---|---|---|---|---|
| It was a disturbance | −1.20 | 0.37 | 1.97 | 2.30 | .32 | 1 |
| It interrupted me in my daily activities | __[a] | — | — | — | — | — |
| It was too time consuming | −1.03 | 0.36 | 0.97 | 2.33 | .31 | 1 |
| The introduction of how to use it was sufficient | −0.53 | 0.38 | 1.46 | 0.87 | .65 | 2 |
| It was boring to use | −0.44 | 0.30 | 2.52 | 2.96 | .23 | 1 |
| It was easy to use | −0.37 | 0.37 | −0.90 | 1.58 | .45 | 2 |
| It was good to use | −0.08 | 0.45 | −1.38 | 2.86 | .24 | 2 |
| The time spent using it has been acceptable | 0.09 | 0.41 | −0.77 | 0.66 | .72 | 2 |
| I can recommend it to others | 0.38 | 0.37 | −0.82 | 1.98 | .37 | 2 |
| It has been difficult to remember to use it | — | — | — | — | — | — |
| It has motivated me to change my lifestyle habits | 0.72 | 0.39 | −2.02 | 3.93 | .14 | 3 |
| It has helped me to understand the benefits of improving my lifestyle habits | 0.73 | 0.40 | −2.44 | 3.01 | .22 | 3 |
| It has helped me to understand how I need to change my lifestyle habits | 0.78 | 0.40 | −2.66 | 1.98 | .37 | 3 |
| It has helped me set personal goals for my lifestyle habits in a way that I could not have done on my own | 0.94 | 0.38 | −1.01 | 1.11 | .57 | 3 |

[a]The items have been removed.

## 12 Item Version

By qualitatively studying the findings and taking the statistics into consideration, it was clear that the items *It has been difficult to remember to use it* and *It interrupted me in my daily activities* did not fit the scale. As a next step, we removed these items and reanalyzed the data. This slightly reduced the targeting (Figure 4) and PSI (from 0.79 to 0.78), but at the same time improved the fit statistics (Table 2). On the basis of the results from the analyses, an updated version, version 2, of our mHealth Satisfaction Questionnaire is presented in Multimedia Appendix 2.

**Figure 4.** Person-item threshold histograms for the 12-item version. Upper histogram (pink bars) shows person measurements reflecting lower leniency with the Health Integrator app to the left, and higher leniency with the Health Integrator app to the right, that is, the most lenient persons are to the right. The lower histogram (blue bars) shows item threshold estimates reflecting lower quality demands to the left and higher quality demands to the right. This implies that it was easier to agree to statements at the lower end corresponding to negative experiences of using the mobile health (mHealth) app compared with the less easy items to agree with at the upper end, corresponding to lifestyle consequences of using the mHealth app.




As shown in Figure 4, there is a gap in items for quality between 0 and 1 logits for the threshold distribution (lower histogram, blue), which could lead to poorer measurement accuracy where most of the respondents are located (upper histogram, pink). There also seems to be room toward the upper end of the scale to include items making more demands for quality using the mHealth app items, and similarly, some less demanding items at the lower end of the quality scale.

Two items showed fit residuals just outside the range −2.5 to +2.5 (*It has helped me to understand how I need to change my lifestyle habits* −2.66; and *It was boring to use* 2.51; Table 2). However, these items showed neither statistically significant $\chi^2$ nor deviating dots from the ICC. Again, a high number of local dependencies (13 of 66 correlations above the relative cut off 0.13) was found, with similar patterns as reported above, and 16.1% (18/112) of the respondents were still outside the desired range of −1.96 to 1.96 when examining unidimensionality. To deal with local dependency, the clustered items were grouped into three testlets (Table 2), and the analyses were repeated. This, however, resulted in disordered thresholds for all testlets. This could not be solved without affecting the content and the satisfactory fit statistics reported above.

## Discussion

### Principal Findings

We present the initial development of a questionnaire, the mHealth Satisfaction Questionnaire, for assessing user satisfaction with mHealth apps and demonstrate a metrological way of evaluating and redesigning the questionnaire's ability to assess this. By applying Rasch analysis, we can better understand the limitations of the questionnaire, as well get guidance about how to revise and improve the questionnaire.

Disordered thresholds indicate that respondents had difficulty in discriminating between the given response options [21]. Unsatisfactory statistics for the monotonicity of item response categories might, however, be a consequence of the small sample size [22]. We solved this by collapsing the 5-point Likert scale into a 3-point Likert scale. With too many response alternatives, there is always a possibility of collapsing response options, but the other way around, that is, splitting responses into two or more categories, cannot be done.

By studying the clusters from the residual correlations used for testlets, three underlying dimensions emerged representing negative experiences of using the mHealth app (testlet 1), positive experiences of using the mHealth app (testlet 2), and lifestyle consequences of using the mHealth app (testlet 3). Given the lack of unidimensionality, it may be questionable to create a single score for a higher ordered assessment of satisfaction with mHealth apps. On the other hand, having a too hardline data–driven approach is not without risk [23]. By considering the clusters that emerged, it was clarified that the items were ordered in a logical hierarchy from the easiest demands for product quality to the more demanding qualities of the product. For mHealth apps, this means that the hierarchy is going from not having negative effects, through positive experiences in the everyday usage, to having a positive impact on lifestyle. Consequently, with those distinct steps in the hierarchy, it can be of importance to specify thresholds or requirement for different types of mHealth apps or to provide guidance for mHealth app developers on what actions are necessary to improve user satisfaction or to ensure higher user satisfaction.

Our analysis showed slightly off-centered persons to the items and that there were some gaps in both item locations and threshold locations. Together with a reliability of 0.78, this

implies that the persons' leniency values are measured with a low precision. This indicates a need for the inclusion of additional items to close the gaps and create a more granulated measure with less measurement uncertainties [12,24]. Adding more items might also help to remove local dependency, as assessments of local dependency seem to be less reliable when there are fewer than 20 items [18]. Future studies exploring which additional items could be brought in, perhaps through respondent interviews to ensure the content validity, may be of value in making sure that the questionnaire captures the full range of concepts of interest.

### Limitations

There are some methodological limitations to bear in mind when interpreting the results. Despite a response rate from our randomized controlled trial of over 80%, the sample size (n=112) could still be considered small. To reach a reliability of 0.8, which is considered an acceptable metrological convention for measurement uncertainties [15], another 14 respondents would have been needed, according to the Spearman-Brown prediction formula [25]. However, in the early stages of methodological work, small sample sizes could be considered acceptable as *convenience samples* for explorative purposes [22].

Another limitation is the frame of reference, that is, that the measurement properties of the mHealth Satisfaction Questionnaire are only evaluated when applied to one unique mHealth app and in one context. However, this is a first step in developing and assessing the measurement properties of the mHealth Satisfaction Questionnaire. With our promising results and suggested improvements for future work, we would recommend that the measurement properties of a future refined version of the mHealth Satisfaction Questionnaire are evaluated further. In particular, DIF, which did not vary with age, gender, or intervention in our sample, would be of interest to study in groups that have used different mHealth apps.

### Comparison With Prior Work

Already in 1986, long before eHealth and later on mHealth had made their entries, Nicell et al [26] designed a questionnaire to measure attitudes toward computers. It included 20 statements like *I feel intimidated by computers* and *Computers are bringing us to a bright new era* with response alternatives on a 5-point Likert scale. Usability, including efficiency, efficacy, and satisfaction, is today a widely accepted metric in many fields and is the focus of several standards and regulations. Initially applied to visual display terminals [27], these standards are now slowly being introduced into medical device and user interface regulations [28,29] and are regulated and promoted by the US Food and Drug Administration [30].

Despite the fact that computers, eHealth, and mHealth have become such an integrated part of everyday life and that mHealth apps commonly act as medical devices, surprisingly little research has been conducted into finding valid and reliable ways to assess user satisfaction with these digital solutions. Besides, different rating scales are increasingly used as outcome measures in clinical studies. Our proposed mHealth Satisfaction Questionnaire, or similar scales evaluated in rigorous ways, can facilitate and guide future development of mHealth solutions and be included in comprehensive usability tests. In addition, such questionnaires may even be of great importance for the accreditation process of mHealth apps and medical technology products.

Our mHealth Satisfaction Questionnaire is an adapted version of similar questionnaires found in the mHealth literature [3,4]. However, how the first versions were developed measurement, and psychometric properties of the previous questionnaires evaluated are not described. They are also more context specific than generic. On the other hand, the mHealth Satisfaction Questionnaire presented in this study provides both an evaluation of its measurement properties and a questionnaire for generic usage in mHealth solutions.

The RMT is considered conceptually and theoretically preferable compared with classical test theory (CTT), both in designing and evaluating rating scales. Limitations with CTT include that (1) data generated are ordinal, (2) scores are scale dependent, (3) scale properties are sample dependent, and (4) data are only suitable for group studies [31]. On the other hand, RMT provides separate estimates of person and item attribute values and their scaling on a common interval logit scale. Moreover, there is a growing interest in RMT in the health care literature [32]. For instance, it has been used to compare measurement performance of questionnaires in diverse areas such as depression in a sample with diverse severity of emotional distress [24], physical and psychological impact of multiple sclerosis [9], and quality of life in sarcoidosis [33], but to our knowledge, psychometric evaluations of rating scales for satisfaction assessment with mHealth have never been conducted. Previously, the Rasch approach to evaluating full usability has only been applied on a few occasions, including in the analysis of Web usability [34] and incontinence product usability [35].

### Conclusions

Taken together, although there is room for improvement, our mHealth Satisfaction Questionnaire gives a new possibility to measure user satisfaction with mHealth. In times where mHealth and digital solutions are given more attention, the mHealth Satisfaction Questionnaire could be an important piece to ensure usability assessments and improve development of mHealth solutions. This paper provides the initial work and suggests further development, where additional items are examined, larger samples are used, and the mHealth Satisfaction Questionnaire is tested for other eHealth apps and in other contexts of use.

### Acknowledgments

The authors would like to acknowledge their health coaches Tove Högman and Carolina Wierbecki for their hard work and dedication in the data collection phase. Furthermore, the authors would like to thank Ellen Franzén, Olov Ståhl, and Pär Hansson

for their thorough work in developing the Health Integrator mHealth app, Petra Sommarlund for excellent European Institute of Innovation and Technology coordination, and Fredrik Söder for his devotion to find collaborators and service offers to the Health Integrator. The Health Integrator was developed as a part of the European collaboration “The health movement—for patient empowerment” (work package ID A1803) supported by the European Institute of Innovation and Technology (grant number 17088). SB also received funding from the Strategic Research Area in Care Sciences (SFO-V), Karolinska Institutet. The funders had no role in the design of the study nor did they have any role in the data collection, analysis, interpretation of data, or in writing of the manuscript.

## Authors' Contributions

YL, SB, and JM designed the study. YL is the primary investigator responsible for the Health Integrator Study. SB is responsible for the data collection, and JL is responsible for the data analyses. JM and YL led the manuscript writing. LP has assisted in analyses. All authors took part in writing of the manuscript and read and approved the final version.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
The mHealth Satisfaction Questionnaire, version 1.
[DOCX File , 68 KB - mhealth_v8i5e15909_app1.docx ]

Multimedia Appendix 2
The mHealth Satisfaction Questionnaire, version 2.
[DOCX File , 41 KB - mhealth_v8i5e15909_app2.docx ]

## References

1. World Health Organization. mHealth: New Horizons for Health Through Mobile Technologies. Geneva: World Health Organization; 2011.
2. Druce KL, Dixon WG, McBeth J. Maximizing engagement in mobile health studies: lessons learned and future directions. Rheum Dis Clin North Am 2019 May;45(2):159-172 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.004] [Medline: 30952390]
3. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS One 2013;8(4):e62613 [FREE Full text] [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
4. Mummah SA, Mathur M, King AC, Gardner CD, Sutton S. Mobile technology for vegetable consumption: a randomized controlled pilot study in overweight adults. JMIR Mhealth Uhealth 2016 May 18;4(2):e51 [FREE Full text] [doi: 10.2196/mhealth.5146] [Medline: 27193036]
5. Bonn SE, Löf M, Östenson CG, Lagerros YT. App-technology to improve lifestyle behaviors among working adults - the Health Integrator study, a randomized controlled trial. BMC Public Health 2019 Mar 7;19(1):273 [FREE Full text] [doi: 10.1186/s12889-019-6595-6] [Medline: 30845949]
6. Pendrill LR. Quality Assured Measurement: Unification Across Social and Physical Sciences. Springer, Cham: Springer; 2019.
7. Wright BD, Linacre JM. Observations are always ordinal; measurements, however, must be interval. Arch Phys Med Rehabil 1989 Nov;70(12):857-860. [Medline: 2818162]
8. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. Chicago: MESA Press; 1993.
9. Pendrill L. Man as a measurement instrument. NCSLI Meas 2014;9(4):24-35. [doi: 10.1080/19315775.2014.11721702]
10. Cano SJ, Pendrill LR, Barbic SP, Fisher WP. Patient-Centred Outcome Metrology for Healthcare Decision-Making. In: Proceedings of the Joint IMEKO TC1-TC7-TC13 Symposium: Measurement Science Challenges in Natural and Social Sciences. 2017 Presented at: IMEKO'17; July 31-August 3, 2017; Rio de Janeiro, Brazil. [doi: 10.1088/1742-6596/1044/1/012057]
11. Andrich D, Sheridan B, Luo G. Interpreting RUMM2030. Perth: RUMM Laboratory Pty Ltd; 2009.
12. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009 Feb;13(12):iii, ix-iiix, 1 [FREE Full text] [doi: 10.3310/hta13120] [Medline: 19216837]
13. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? Arthritis Rheum 2007 Dec 15;57(8):1358-1362 [FREE Full text] [doi: 10.1002/art.23108] [Medline: 18050173]
14. Nunnally JC, Bernstein IH. Psychometric Theory. Third Edition. New York, New York: McGraw-Hill; 1994.
15. Pendrill LR, Fisher WP. Counting and quantification: comparing psychometric and metrological perspectives on visual perceptions of number. Meas 2015 Jul;71:46-55 [FREE Full text] [doi: 10.1016/j.measurement.2015.04.010]

16. Hagquist C, Andrich D. Recent advances in analysis of differential item functioning in health research using the Rasch model. Health Qual Life Outcomes 2017 Sep 19;15(1):181 [FREE Full text] [doi: 10.1186/s12955-017-0755-0] [Medline: 28927468]
17. Christensen KB, Makransky G, Horton M. Critical values for Yen's : identification of local dependence in the Rasch model using residual correlations. Appl Psychol Meas 2017 May;41(3):178-194 [FREE Full text] [doi: 10.1177/0146621616677520] [Medline: 29881087]
18. Marais I. Local dependence. In: Christensen KB, Mesbah M, Kreiner S, editors. Rasch Models in Health. London, UK: John Wiley & Sons, Inc; 2013.
19. Marais I, Andrich D. Formalizing dimension and response violations of local independence in the unidimensional Rasch model. J Appl Meas 2008;9(3):200-215. [Medline: 18753691]
20. Smith RM. A comparison of methods for determining dimensionality in Rasch measurement. Struct Equ Modeling 1996 Jan;3(1):25-40. [doi: 10.1080/10705519609540027]
21. Andrich D. A rating formulation for ordered response categories. Psychometrika 1978 Dec;43(4):561-573. [doi: 10.1007/bf02293814]
22. Chen W, Lenderking W, Jin Y, Wyrwich KW, Gelhorn H, Revicki DA. Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. Qual Life Res 2014 Mar;23(2):485-493. [doi: 10.1007/s11136-013-0487-5] [Medline: 23912855]
23. Morel T, Cano SJ. Measuring what matters to rare disease patients - reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis 2017 Nov 2;12(1):171 [FREE Full text] [doi: 10.1186/s13023-017-0718-x] [Medline: 29096663]
24. Cleanthous S, Barbic SP, Smith S, Regnault A. Psychometric performance of the PROMIS depression item bank: a comparison of the 28- and 51-item versions using Rasch measurement theory. J Patient Rep Outcomes 2019 Jul 30;3(1):47 [FREE Full text] [doi: 10.1186/s41687-019-0131-4] [Medline: 31359206]
25. Spearman C. Correlation calculated from faulty data. Br J Psychol 1910;3(3):271-295. [doi: 10.1111/j.2044-8295.1910.tb00206.x]
26. Nickell GS, Pinto JN. The computer attitude scale. Comput Hum Behav 1986 Jan;2(4):301-306. [doi: 10.1016/0747-5632(86)90010-5]
27. International Organization for Standardization. ISO 9241-11:1998 Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs) — Part 11: Guidance on Usability URL: https://www.iso.org/standard/16883.html [accessed 2019-08-01]
28. International Organization for Standardization. IEC 62366-1:2015 Medical Devices — Part 1: Application of Usability Engineering to Medical Devices URL: https://www.iso.org/standard/63179.html [accessed 2019-08-01]
29. The Government of Sweden. 2011. National eHealth - The Strategy for Accessible and Secure Information in Health and Social Care URL: https://www.government.se/49b757/contentassets/65d092eda66b437e9cb5e54bb4e41319/national-ehealth---the-strategy-for-accessible-and-secure-information-in-health-and-social-care-s2011.023 [accessed 2019-08-01]
30. Food and Drug Administration. 2016. Applying Human Factors and Usability Engineering to Medical Devices URL: https://www.fda.gov/media/80481/download [accessed 2019-08-01]
31. Cano S, Hobart J. The problem with health measurement. Patient Prefer Adherence 2011 Jun;2011(5):279-290. [doi: 10.2147/ppa.s14399]
32. Leung Y, Png M, Conaghan P, Tennant A. A systematic literature review on the application of Rasch analysis in musculoskeletal disease -- a special interest group report of OMERACT 11. J Rheumatol 2014 Jan;41(1):159-164. [doi: 10.3899/jrheum.130814] [Medline: 24128781]
33. Patel AS, Siegert RJ, Creamer D, Larkin G, Maher TM, Renzoni EA, et al. The development and validation of the King's Sarcoidosis Questionnaire for the assessment of health status. Thorax 2013 Jan;68(1):57-65. [doi: 10.1136/thoraxjnl-2012-201962] [Medline: 23065052]
34. Tezza R, Bornia AC, de Andrade DF. Measuring web usability using item response theory: principles, features and opportunities. Interact Comput 2011 Mar;23(2):167-175. [doi: 10.1016/j.intcom.2011.02.004]
35. Rice SK, Pendrill L, Petersson N, Nordlinder J, Farbrot A. Rationale and design of a novel method to assess the usability of body-worn absorbent incontinence care products by caregivers. J Wound Ostomy Continence Nurs 2018;45(5):456-464 [FREE Full text] [doi: 10.1097/WON.0000000000000462] [Medline: 30188394]

## Abbreviations

**CTT:** classical test theory
**DIF:** differential item functioning
**eHealth:** electronic health
**ICC:** item characteristic curve
**mHealth:** mobile health

**PSI:** Person Separation Index
**RMT:** Rasch measurement theory

*Edited by G Eysenbach; submitted 17.08.19; peer-reviewed by B Chaudhry, Y Wang; comments to author 09.12.19; revised version received 20.01.20; accepted 07.02.20; published 26.05.20.*

*Please cite as:*
*Melin J, Bonn SE, Pendrill L, Trolle Lagerros Y*
*A Questionnaire for Assessing User Satisfaction With Mobile Health Apps: Development Using Rasch Measurement Theory*
*JMIR Mhealth Uhealth 2020;8(5):e15909*
*URL: http://mhealth.jmir.org/2020/5/e15909/*
*doi:10.2196/15909*
*PMID:32452817*

Original Paper

# Wrist-Worn Wearables for Monitoring Heart Rate and Energy Expenditure While Sitting or Performing Light-to-Vigorous Physical Activity: Validation Study

Peter Düking[1*], MSc; Laura Giessing[2*], MSc; Marie Ottilie Frenkel[2*], PhD; Karsten Koehler[3*], PhD; Hans-Christer Holmberg[4,5*], PhD; Billy Sperlich[1*], PhD

[1]Integrative and Experimental Exercise Science, Department of Sport Science, University of Würzburg, Würzburg, Germany

[2]Department of Sport Psychology, Institute for Sport and Sport Sciences, Heidelberg University, Heidelberg, Germany

[3]Department of Sport and Health Science, Technical University of Munich, Munich, Germany

[4]Department of Health Sciences, Mid Sweden University, Östersund, Sweden

[5]Department of Physiology and Pharmacology, Karolinska Institute, Stockholm, Sweden

[*]all authors contributed equally

**Corresponding Author:**
Peter Düking, MSc
Integrative and Experimental Exercise Science
Department of Sport Science
University of Würzburg
Judenbühlweg 11
Würzburg, 97082
Germany
Phone: 49 931 31 ext 8479
Email: peterdueking@gmx.de

## Abstract

**Background:** Physical activity reduces the incidences of noncommunicable diseases, obesity, and mortality, but an inactive lifestyle is becoming increasingly common. Innovative approaches to monitor and promote physical activity are warranted. While individual monitoring of physical activity aids in the design of effective interventions to enhance physical activity, a basic prerequisite is that the monitoring devices exhibit high validity.

**Objective:** Our goal was to assess the validity of monitoring heart rate (HR) and energy expenditure (EE) while sitting or performing light-to-vigorous physical activity with 4 popular wrist-worn wearables (Apple Watch Series 4, Polar Vantage V, Garmin Fenix 5, and Fitbit Versa).

**Methods:** While wearing the 4 different wearables, 25 individuals performed 5 minutes each of sitting, walking, and running at different velocities (ie, 1.1 m/s, 1.9 m/s, 2.7 m/s, 3.6 m/s, and 4.1 m/s), as well as intermittent sprints. HR and EE were compared to common criterion measures: Polar-H7 chest belt for HR and indirect calorimetry for EE.

**Results:** While monitoring HR at different exercise intensities, the standardized typical errors of the estimates were 0.09-0.62, 0.13-0.88, 0.62-1.24, and 0.47-1.94 for the Apple Watch Series 4, Polar Vantage V, Garmin Fenix 5, and Fitbit Versa, respectively. Depending on exercise intensity, the corresponding coefficients of variation were 0.9%-4.3%, 2.2%-6.7%, 2.9%-9.2%, and 4.1%-19.1%, respectively, for the 4 wearables. While monitoring EE at different exercise intensities, the standardized typical errors of the estimates were 0.34-1.84, 0.32-1.33, 0.46-4.86, and 0.41-1.65 for the Apple Watch Series 4, Polar Vantage V, Garmin Fenix 5, and Fitbit Versa, respectively. Depending on exercise intensity, the corresponding coefficients of variation were 13.5%-27.1%, 16.3%-28.0%, 15.9%-34.5%, and 8.0%-32.3%, respectively.

**Conclusions:** The Apple Watch Series 4 provides the highest validity (ie, smallest error rates) when measuring HR while sitting or performing light-to-vigorous physical activity, followed by the Polar Vantage V, Garmin Fenix 5, and Fitbit Versa, in that order. The Apple Watch Series 4 and Polar Vantage V are suitable for valid HR measurements at the intensities tested, but HR data provided by the Garmin Fenix 5 and Fitbit Versa should be interpreted with caution due to higher error rates at certain intensities. None of the 4 wrist-worn wearables should be employed to monitor EE at the intensities and durations tested.

*(JMIR Mhealth Uhealth 2020;8(5):e16716)* doi:10.2196/16716

**KEYWORDS**

cardiorespiratory fitness; innovation; smartwatch; technology; wearable; digital health

## Introduction

Physical activity reduces the incidences of noncommunicable diseases, obesity, and mortality, but, unfortunately, according to the World Health Organization (WHO), a sedentary lifestyle is becoming increasingly common, with approximately 23% of the adult population failing to meet physical activity guidelines [1-3]. Accordingly, innovative approaches to promote and monitor physical activity are urgently warranted, as indicated in the WHO's global action plan [4]. While individual monitoring of physical activity aids in the design of effective interventions to enhance physical activity [5,6], a basic prerequisite is that the monitoring devices exhibit high validity.

Heart rate (HR) and energy expenditure (EE) are two key aspects of physical activity. HR reflects the intensity of physical activity [7,8], while monitoring EE is particularly helpful for individuals seeking to regulate their body mass or composition [9], since any imbalance between energy intake and EE may have negative consequences [10]. HR and EE vary widely between individuals, and careful monitoring is crucial to provide appropriate recommendations concerning physical activity and diet [10].

While several procedures for monitoring HR (eg, Holter monitors or chest belts) and EE (indirect calorimetry) are available, miniaturized sensors [11] potentially enable less restrictive monitoring. Utilization of data collected by miniaturized wearable sensors (wearables) to improve health and fitness is a current worldwide trend [12] that offers new opportunities for designing individualized interventions concerning physical activity [13]. Theoretically, wearables allow extensive monitoring of parameters related to physical activity over prolonged periods [14]. Rigorous validation of wearable sensors is paramount since insurance companies encourage and promote monitoring (with wearables representing a major component of this strategy) [15], the WHO aims to endorse digital health (including wearables) [16], and in Germany, state laws already permit physicians to prescribe digital health solutions [17].

Wearable manufacturers claim to enable noninvasive and accurate monitoring of HR and EE [18]. The market for wearables designed to improve health and fitness is growing rapidly, and companies release new versions of their technology at least once each year, with older versions disappearing from the market. Projections for wrist-worn wearables alone estimate that 152.7 million such devices will be shipped in 2019, with a compound annual growth rate of 6.2% until 2023 [19]. However, the validity of most commercially available wearables has not been assessed across a range of exercise intensities by independent research institutions [18,20,21]. Consequently, while the potential health benefits of wearables are considerable, their validity must first be assured.

Accordingly, the current investigation was designed to assess the validity of 4 commercially available**,** high-tech, and popular wearable models for monitoring HR and EE while sitting or performing light-to-vigorous physical exercise.

## Methods

Our study protocol and data analysis were based on previous recommendations concerning the validation of the reliability of wearables for assessing parameters during physical activity [22].

### Participants

After being informed about the experimental procedures, 25 healthy participants (11 men, 14 women; mean age 26 years, SD 7 years; mean body height 174 cm, SD 10 cm; mean body mass 70.1 kg, SD 12.0 kg) of Caucasian origin gave their written consent to participate. This study was performed in accordance with the Declaration of Helsinki and approved by our institute's ethical committee (Ethical approval number: EthikKomm-05/2019).

### Experimental Procedures

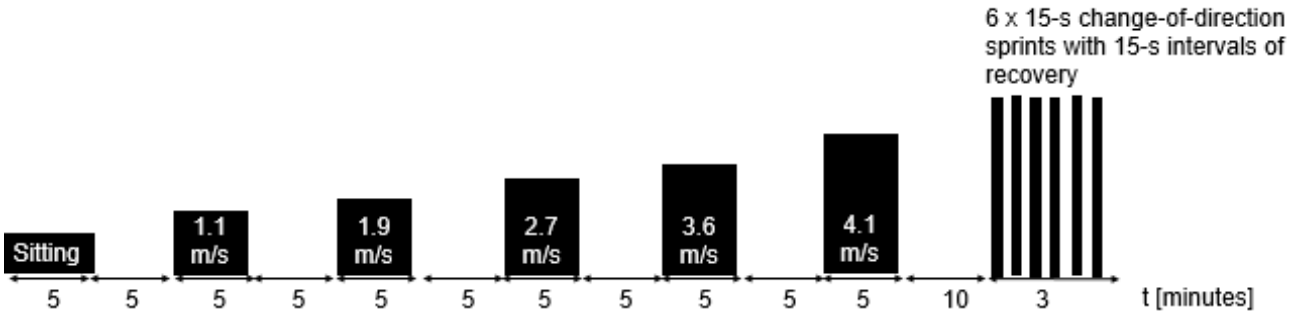
All participants visited the laboratory twice, with 3 days between visits, and tested 2 different wearables on each occasion. Environmental conditions were constant, with a temperature of 19.5 °C (SD 0.8 °C). Anthropometric data were collected during the first visit. Each wearable was attached to the wrist in the manner indicated by the manufacturer, and age, sex, height, and body mass were entered into the wearable's software, along with information about whether the wearable was on the left or right wrist.

The wearables and the order in which they were worn during the first and second visits were chosen in a random fashion, resulting in 25 measurements with each wearable.

Each participant was monitored while sitting as well as during walking and running at different speeds (1.1 m/s, 1.9 m/s, 2.7 m/s, 3.6 m/s, and 4.1 m/s) for 5 minutes, interspersed with 5 minutes of standing still. All participants also performed 6 ~30-m sprints involving multiple changes in direction (ranging from 10° to 180°) on the SpeedCourt (GlobalSpeed GmbH, Hemsbach, Germany) [23]. This involved sprinting between 12 contact plates installed symmetrically in a 5.25 m by 5.25 m square on the floor. A software program designed a path consisting of the 6 30-m sprints (approximately 15 seconds per 30-m sprint), with a display indicating the contact plates that had to be touched [23].

Figure 1 summarizes the sitting, walking, and running procedures.

**Figure 1.** Schematic illustration of the periods during which each participant was monitored (black bars).




### *Criterion Measures*

A portable breath-by-breath gas analyzer (Metamax 3B, CORTEX Biophysik GmbH, Leipzig, Germany) employing standard algorithms for indirect calorimetry served as the criterion measure for EE. This system measures metabolic demands reliably [24] and has been used previously to assess the validity of wearables designed to monitor EE [25].

A Polar H7 chest belt, commonly employed for similar evaluations [26,27], was synchronized with the gas analyzer and served as the criterion measure for HR.

### *Wearables*

The 4 tested wrist-worn wearables were Apple Watch Series 4, Version 5.1 (Apple Inc, Cupertino, CA); Polar Vantage V, Firmware 3.1.7 (Polar Electro Oy, Kempele, Finland); Garmin Fenix 5, Software 7.6 (Garmin, Olathe, KS); and Fitbit Versa, Version 32.33.1.30 (Fitbit Inc, San Francisco, CA).

All utilize photoplethysmography to monitor HR, but, to the best of our knowledge, information concerning the data used to calculate EE is not publicly available. Each wearable was positioned firmly, yet comfortably, on the wrist as in real life and as recommended by the manufacturers.

In the case of the Apple Watch Series 4, the “indoor walking” mode was selected for measurements while sitting or walking at 1.1 m/s; “running indoor” for speeds from 1.9 m/s to 4.1 m/s; and “HIIT” for the intermittent sprints. For the Polar Vantage V, the “Running (Treadmill)” mode was selected for all the monitoring periods, except for the intermittent sprints involving many and frequent changes in direction, for which “Soccer” was chosen. With the Garmin Fenix 5 and Fitbit Versa, the “Treadmill” mode was chosen for all monitoring periods.

All data were transmitted via Bluetooth and synchronized with the accompanying smartphone applications, in accordance with the manufacturers’ recommendations. For the Apple Watch Series 4, the raw data were exported to Microsoft Excel (Microsoft Corp, Redmond, WA) via the Apple Health App (Apple Inc, Cupertino, CA). In the cases of Polar, Garmin, and Fitbit, data were exported via specific buttons in the accompanying online software or collected directly from the software.

### Statistical Analysis

Statistical analysis was performed in accordance with previous recommendations, whenever applicable [22]. Prior to analysis, the data were log-transformed to avoid bias resulting from nonuniformity of error. All data were analyzed in custom-designed Microsoft Excel spreadsheets [28]. For each exercise, the standardized mean bias was calculated. As recommended and carried out previously, linear regression was employed to analyze validity [22,29]. The standardized mean bias, standardized typical error of the estimate (sTEE), coefficient of variation (CV), and Pearson’s product-moment correlation coefficient are all reported.

The sTEE, based on half the thresholds of the modified Cohen’s scale, was employed to assess validity: <0.1, trivial; 0.1-0.29, small; 0.3-0.59, moderate; 0.6-1.0, large; 1.0-2.0, very large; >2.0, extremely large [28]. Pearson’s *r* was utilized to evaluate the correlation between the criterion measure and wearable as follows: 0.45-0.69, very poor; 0.70-0.84, poor; 0.85-0.94, good; 0.95-0.994, very good; ≥0.995, excellent [30]. The 90% confidence limits (coefficient of variation [CV]) for the statistical parameters are also reported. Absolute errors were calculated based on these CVs and the mean value obtained by the criterion measure.

The level of physical activity was defined in terms of the metabolic equivalent (MET), with <3 MET indicating light, <6 MET medium, and >6 MET vigorous physical activity [31]. To define physical activity levels, the EE provided by the criterion measure was extrapolated to 1 hour and divided by the mean body weight of the participant.

## Results

### Heart rate

The mean HR, CV, Pearson’s *r*, and sTEE with 90% confidence limits and interpretations are summarized in Table 1.

**Table 1.** Analysis of the validity of heart rate measurements by wrist-worn wearables while sitting or walking/running at different intensities.

| Level of activity (METs[a]), intensity | Apple Watch Series 4 | Polar Vantage V | Garmin Fenix 5 | Fitbit Versa |
|---|---|---|---|---|
| **Inactive (1.3), sitting** | | | | |
| Heart rate (bpm)[b], mean (SD) | 68.8 (11.7) | | | |
| Standardized mean bias | 0.03 (–0.02 to 0.07) | –0.06 (–0.11 to –0.02) | 0.12 (–0.07 to 0.31) | –0.06 (–0.27 to 0.15) |
| Pearson's *r* | 0.99 (0.99-1) | 0.99 (0.98-1) | 0.89 (0.77-0.95) | 0.91 (0.77-0.96) |
| Interpretation of Pearson's *r* | Excellent | Excellent | Good | Good |
| CV[c] (%) | 2 (1.6-2.6) | 2.2 (1.8-2.9) | 7.7 (6.1-10.7) | 8 (6.1-12.1) |
| sTEE[d] | 0.12 (0.09-0.17) | 0.13 (0.10-0.19) | 0.63 (0.41-1.03) | 0.47 (0.28-0.82) |
| Interpretation of sTEE | Small | Small | Large | Moderate |
| **Light (3.5), 1.1 m/s** | | | | |
| Heart rate (bpm)[b], mean (SD) | 95.8 (25.0) | | | |
| Standardized mean bias | 0.01 (–0.07 to 0.09) | –0.07 (–0.32 to 0.17) | 0.12 (–0.10 to 0.34) | –0.28 (–7.00 to 0.13) |
| Pearson's *r* | 0.97 (0.95-0.99) | 0.89 (0.79-0.94) | 0.85 (0.70-0.93) | 0.57 (0.31-0.70) |
| Interpretation of Pearson's *r* | Very good | Good | Good | Very poor |
| CV (%) | 2.9 (2.3-3.8) | 5.5 (4.4-7.3) | 5.8 (4.5-8.0) | 9.6 (7.8-12.6) |
| sTEE | 0.23 (0.16-0.34) | 0.54 (0.37-0.82) | 0.62 (0.40-1.03) | 1.43 (0.87-3.03) |
| Interpretation of sTEE | Small | Moderate | Large | Very large |
| **Vigorous (6.6), 1.9 m/s** | | | | |
| Heart rate (bpm)[b], mean (SD) | 127 (19.4) | | | |
| Standardized mean bias | –0.02 (–0.10 to 0.06) | –0.34 (–0.53 to –0.16) | 0.06 (–0.17 to 0.29) | –0.05 (–0.34 to 0.24) |
| Pearson's *r* | 0.97 (0.95-0.99) | 0.91 (0.82-0.95) | 0.83 (0.65-0.92) | 0.54 (0.29-0.71) |
| Interpretation of Pearson's *r* | Very good | Good | Poor | Very poor |
| CV (%) | 2.9 (2.3-3.8) | 5.4 (4.3-7.2) | 9.2 (7.2-12.9) | 19.1 (15.7-24.7) |
| sTEE | 0.23 (0.16-0.34) | 0.46 (0.32-0.69) | 0.68 (0.43-1.16) | 1.58 (0.98-3.25) |
| Interpretation of sTEE | Small | Moderate | Large | Very large |
| **Vigorous (9.9), 2.7 m/s** | | | | |
| Heart rate (bpm)[b], mean (SD) | 167 (16.5) | | | |
| Standardized mean bias | –0.13 (–0.49 to 0.24) | –0.37 (–0.57 to –0.16) | –0.56 (–0.87 to –0.24) | –0.82 (–1.18 to –0.47) |
| Pearson's *r* | 1 (0.99-1) | 0.88 (0.78-0.94) | 0.63 (0.34-0.81) | 0.52 (0.27-0.70) |
| Interpretation of Pearson's *r* | Excellent | Good | Very poor | Very poor |
| CV (%) | 0.9 (0.7-1.2) | 5.9 (4.8-7.9) | 8.3 (6.6-11.4) | 8.5 (7.0-11.0) |
| sTEE | 0.09 (0.06-0.12) | 0.53 (0.36-0.81) | 1.24 (0.74-2.73) | 1.64 (1.01-3.59) |
| Interpretation of sTEE | Trivial | Moderate | Very large | Very large |
| **Vigorous (10.4), 3.6 m/s** | | | | |
| Heart rate (bpm)[b], mean (SD) | 170 (15.3) | | | |
| Standardized mean bias | 0.02 (–0.09 to 0.14) | –0.75 (–1.05 to –0.46) | –0.40 (–0.60 to –0.19) | –1.17 (–1.47 to –0.87) |
| Pearson's *r* | 0.94 (0.89-0.97) | 0.86 (0.74-0.93) | 0.82 (0.67-0.91) | 0.82 (0.67-0.91) |
| Interpretation of Pearson's *r* | Good | Good | Poor | Poor |
| CV (%) | 3.0 (2.4-4.0) | 4.9 (3.9-6.5) | 8.9 (7.19-12.1) | 4.1 (3.3-5.5) |
| sTEE | 0.35 (0.24-0.51) | 0.59 (0.40-0.91) | 0.69 (0.46-1.11) | 0.70 (0.47-1.11) |
| Interpretation of sTEE | Moderate | Moderate | Large | Large |

| Level of activity (METs[a]), intensity | Apple Watch Series 4 | Polar Vantage V | Garmin Fenix 5 | Fitbit Versa |
|---|---|---|---|---|
| **Vigorous (13.3), 4.1 m/s** | | | | |
| Heart rate (bpm)[b], mean (SD) | 177 (8.5) | | | |
| Standardized mean bias | –0.27 (–0.51 to –0.03) | –0.72 (–0.95 to –0.49) | –1.47 (–1.88 to –1.06) | –2.06 (–3.17 to –0.95) |
| Pearson's *r* | 0.85 (0.71-0.93) | 0.89 (0.76-0.95) | 0.82 (0.65-0.91) | 0.68 (0.24-0.89) |
| Interpretation of Pearson's *r* | Good | Good | Poor | Very poor |
| CV (%) | 4.3 (3.4-5.8) | 3.9 (3.0-5.6) | 2.88 (2.28-3.96) | 3.22 (2.34-5.35) |
| sTEE | 0.62 (0.41-1.00) | 0.50 (0.31-0.84) | 0.69 (0.44-1.17) | 1.09 (0.52-4.13) |
| Interpretation of sTEE | Large | Moderate | Large | Very large |
| **Vigorous (13.8), intermittent sprints** | | | | |
| Heart rate (bpm)[b], mean (SD) | 153 (14.7) | | | |
| Standardized mean bias | 0.12 (0.03 to 0.21) | –0.99 (–1.54 to –0.44) | –1.75 (–2.28 to –1.21) | –2.01 (–2.58 to –1.43) |
| Pearson's *r* | 0.92 (0.85-0.96) | 0.75 (0.53-0.88) | 0.58 (0.28-0.78) | 0.53 (0.15-0.77) |
| Interpretation of Pearson's *r* | Good | Poor | Very poor | Very poor |
| CV (%) | 3.5 (2.8-4.7) | 6.7 (5.3-9.3) | 8.4 (6.6-11.6) | 9.0 (6.9-13.4) |
| sTEE | 0.38 (0.25-0.64) | 0.88 (0.54-1.73) | 1.44 (0.80-5.40) | 1.94 (0.84-5.25) |
| Interpretation of sTEE | Moderate | Large | Very large | Very large |
| **Vigorous (8.8), average of the values at all different intensities** | | | | |
| Heart rate (bpm)[b], mean | 137 | | | |
| Standardized mean bias | 0.03 | –0.47 | –0.55 | –0.92 |
| Pearson's *r* | 0.95 | 0.88 | 0.77 | 0.65 |
| Interpretation of Pearson's *r* | Very good | Good | Poor | Very poor |
| CV (%) | 2.79 | 4.93 | 7.30 | 8.79 |
| sTEE | 0.29 | 0.52 | 0.86 | 1.26 |
| Interpretation of sTEE | Moderate | Moderate | Large | Very large |

[a]METs: metabolic equivalents.

[b]Measured according to the criterion measure.

[c]CV: coefficient of variation.

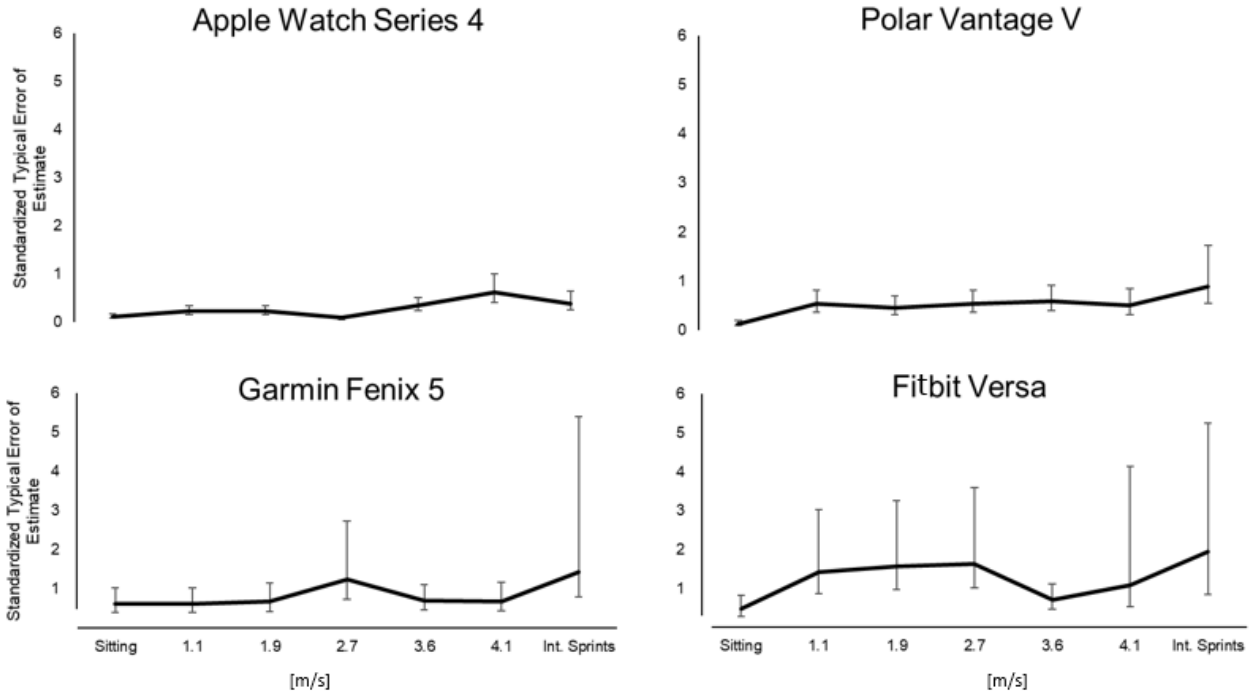
[d]sTEE: standardized typical error of the estimate.

Figure 2 documents the sTEE for the HR values provided by the wearables at all exercise intensities.

For HR monitoring at the different intensities, the sTEE was 0.09-0.62, 0.13-0.88, 0.62-1.24, and 0.47-1.94 for the Apple Watch Series 4, Polar Vantage V, Garmin Fenix 5, and Fitbit Versa, respectively, with corresponding CVs of 0.9%-4.3%, 2.2%-6.7%, 2.88%-9.2%, and 4.1%-19.1%, respectively. The sTEE was less affected by intensity in the case of the Apple Watch Series 4 and Polar Vantage V devices than with the Garmin Fenix 5 and Fitbit Versa devices.

sTEE and CV peaked during the intermittent sprints for all the wearables except the Apple Watch Series 4.

**Figure 2.** Standardized typical errors of the estimate (90% CI) for heart rate monitoring by the wearables while sitting or performing light-to-vigorous physical activity.




## Energy Expenditure

The mean EE, CV, Pearson's correlation coefficient, and sTEE with 90% confidence limits and interpretations are shown in Table 2.

Figure 3 depicts the sTEE for the EE values provided by all 4 wearables during exercise at different intensities.

These sTEE values were 0.34-1.84, 0.32-1.33, 0.46-4.86, and 0.41-1.65 for the Apple Watch Series 4, Polar Vantage V, Garmin Fenix 5, and Fitbit Versa, respectively, with corresponding CVs of 13.5%-27.1%, 16.3%-28.0%, 15.9%-34.5%, and 8.0%-32.3%, respectively.

**Table 2.** Analysis of the validity of energy expenditure measurements by wrist-worn wearables while sitting and walking/running at different intensities.

| Level of activity (METs[a]), intensity | Apple Watch Series 4 | Polar Vantage V | Garmin Fenix 5 | Fitbit Versa |
|---|---|---|---|---|
| **Inactive (1.3), sitting** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 7.6 (1.6) | | | |
| Standardized mean bias | 2.59 (2.25 to 2.94) | 0.25 (–0.40 to 0.90) | 1.74 (0.77 to 2.71) | –0.72 (–1.46 to 0.02) |
| Pearson's *r* | 0.46 (0.16 to 0.68) | 0.41 (0.10 to 0.65) | 0.23 (–0.15 to 0.55) | 0.52 (0.16 to 0.76) |
| Interpretation of Pearson's *r* | Very poor | - | - | Very poor |
| CV[c] (%) | 26.6 (21.2-36.2) | 28.0 (22.2-38.4) | 20.9 (16.3-29.7) | 17.1 (13.2-24.7) |
| sTEE[d] | 1.84 (1.02-5.64) | 1.33 (0.79-2.94) | 4.24 (1.51-6.46) | 1.65 (0.87-6.09) |
| Interpretation of sTEE | Very large | Very large | Extremely large | Very large |
| **Light (3.5), 1.1 m/s** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 20.6 (4.1) | | | |
| Standardized mean bias | 2.63 (2.23 to 2.03 | 1.29 (0.87 to 1.72) | –0.05 (–0.84 to 0.74) | 4.16 (3.97 to 4.36) |
| Pearson's *r* | 0.71 (0.49 to 0.85) | 0.67 (0.44 to 0.82) | 0.20 (–0.19 to 0.54) | 0.88 (0.76 to 0.94) |
| Interpretation of Pearson's *r* | Poor | Very poor | - | Good |
| CV (%) | 15.1 (12.0-20.5) | 16.3 (13.1-22.1) | 16.8 (13.1-24.0) | 8.0 (6.3-11.2) |
| sTEE | 0.99 (0.63-1.77) | 1.10 (0.70-2.03) | 4.86 (1.56-5.11) | 0.53 (0.35-0.85) |
| Interpretation of sTEE | Large | Very large | Extremely large | Moderate |
| **Vigorous (6.6), 1.9 m/s** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 38.3 (6.5) | | | |
| Standardized mean bias | 1.58 (1.27 to 1.90) | 0.27 (–0.18 to 0.71) | –1.15 (–2.01 to –0.29) | 0.88 (0.56 to 1.20) |
| Pearson's *r* | 0.71 (0.49 to 0.84) | 0.49 (0.18 to 0.7) | 0.21 (–0.21 to 0.56) | 0.78 (0.57 to 0.89) |
| Interpretation of Pearson's *r* | Poor | Very poor | - | Poor |
| CV (%) | 13.5 (10.8-18.1) | 17.1 (13.7-23.1) | 15.9 (12.2-23.3) | 11.2 (8.8-15.7) |
| sTEE | 0.99 (0.64-1.76) | 0.65 (0.43-1.02) | 4.62 (1.46-4.73) | 0.81 (0.51-1.44) |
| Interpretation of sTEE | Large | Large | Extremely large | Large |
| **Vigorous (9.9), 2.7 m/s** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 57.8 (11.0) | | | |
| Standardized mean bias | 0.79 (0.56 to 1.02) | –0.09 (–0.39 to 0.2) | –0.04 (–0.45 to 0.37) | –0.06 (–0.44 to 0.32) |
| Pearson's *r* | 0.80 (0.62 to 0.90) | 0.72 (0.51 to 0.85) | 0.57 (0.25 to 0.78) | 0.74 (0.51 to 0.87) |
| Interpretation of Pearson's *r* | Poor | Poor | Very poor | Poor |
| CV (%) | 19.0 (15.1-26.2) | 21.9 (17.5-29.8) | 17.1 (13.3-24.4) | 14.1 (11-19.8) |
| sTEE | 0.76 (0.50-1.25) | 0.97 (0.62-1.68) | 1.43 (0.80-3.91) | 0.90 (0.50-1.67) |
| Interpretation of sTEE | Large | Large | Very large | Large |
| **Vigorous (10.4), 3.6 m/s** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 60.5 (26.7) | | | |
| Standardized mean bias | 0.32 (0.19 to 0.45) | –0.05 (–0.18 to 0.08) | 0.19 (–0.10 to 0.48) | –0.06 (–0.37 to 0.24) |
| Pearson's *r* | 0.95 (0.89 to 0.97) | 0.95 (0.89 to 0.97) | 0.84 (0.68 to 0.92) | 0.76 (0.52 to 0.88) |
| Interpretation of Pearson's *r* | Very good | Very good | Poor | Poor |

| Level of activity (METs[a]), intensity | Apple Watch Series 4 | Polar Vantage V | Garmin Fenix 5 | Fitbit Versa |
|---|---|---|---|---|
| CV (%) | 20.3 (16.0-28.3) | 20.7 (16.4-28.6) | 34.5 (26.4-50.8) | 32.3 (24.6-48) |
| sTEE | 0.34 (0.23-0.50) | 0.34 (0.24-0.51) | 0.64 (0.41-1.09) | 0.87 (0.53-1.65) |
| Interpretation of sTEE | Moderate | Moderate | Large | Large |
| **Vigorous (13.3), 4.1 m/s** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 77.8 (46.6) | | | |
| Standardized mean bias | 0.34 (0.13 to 0.54) | –0.11 (–0.28 to 0.05) | 0.25 (–0.06 to 0.55) | 0.13 (–0.09 to 0.34) |
| Pearson's *r* | 0.93 (0.82 to 0.98) | 0.95 (0.87 to 0.98) | 0.91 (0.78 to 0.96) | 0.92 (0.81 to 0.97) |
| Interpretation of Pearson's *r* | Good | Very good | Good | Good |
| CV (%) | 27.1 (19.6-45.1) | 22.7 (16.5-37.3) | 33.1 (24.3-52.9) | 29.9 (21.8-48.6) |
| sTEE | 0.39 (0.23-0.71) | 0.32 (0.19-0.57) | 0.46 (0.28-0.80) | 0.41 (0.24-0.72) |
| Interpretation of sTEE | Moderate | Moderate | Moderate | Moderate |
| **Vigorous (13.8), intermittent sprints** | | | | |
| Energy expenditure (kcal/5 min)[b], mean (SD) | 80.4 (15.6) | | | |
| Standardized mean bias | 1.83 (1.52 to 2.13) | 0.23 (0.04 to 0.42) | –0.82 (–1.78 to 0.14) | –1.25 (–1.83 to –0.67) |
| Pearson's *r* | 0.66 (0.41 to 0.81) | 0.85 (0.72 to 0.92) | 0.21 (–0.19 to 0.56) | 0.42 (0.06 to 0.68) |
| Interpretation of Pearson's *r* | Very poor | Good | - | - |
| CV (%) | 25.4 (20.2-34.7) | 17.5 (14.0-23.6) | 17.9 (13.8-25.9) | 20.8 (16.2-29.6) |
| sTEE | 1.15 (0.72-2.19) | 0.63 (0.43-0.97) | 4.62 (1.50-5.05) | 1.64 (0.88-5.57) |
| Interpretation of sTEE | Very large | Large | Extremely large | Very large |
| **Vigorous (8.8), average of the values at all different intensities** | | | | |
| Energy expenditure (kcal/5 min)[b], mean | 49.0 | | | |
| Standardized mean bias | 1.44 | 0.26 | 0.02 | 0.44 |
| Pearson's *r* | 0.75 | 0.72 | 0.45 | 0.72 |
| Interpretation of Pearson's *r* | Poor | Poor | Very poor | Poor |
| CV (%) | 21.0 | 20.6 | 22.3 | 19.1 |
| sTEE | 0.92 | 0.76 | 2.98 | 0.97 |
| Interpretation of sTEE | Large | Large | Extremely large | Large |

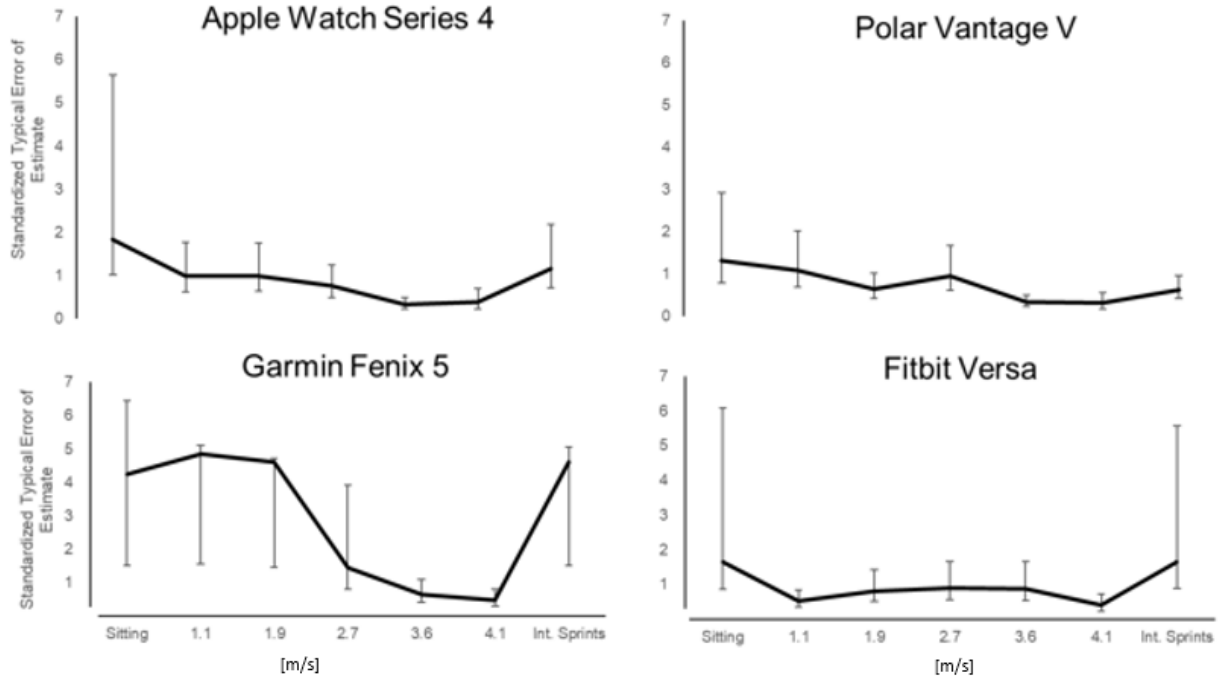
[a]METs: metabolic equivalents.

[b]Measured according to the criterion measure.

[c]CV: coefficient of variation.

[d]sTEE: standardized typical error of the estimate.

**Figure 3.** Standardized typical errors of the estimate (90% CI) for energy expenditure monitoring by the wearables while sitting or performing light-to-vigorous physical activity.




## Discussion

### Principal Findings

The current investigation was designed to assess the validity of 4 commercially available wrist-worn wearables for monitoring HR and EE while sitting or performing light-to-vigorous physical activity.

The following paragraphs outline our major findings.

For monitoring HR during sitting or walking/running up to 2.7 m/s or with a HR up to 167 bpm, the Apple Watch Series 4 demonstrated the highest validity (average 2.3 bpm deviation from the criterion measure), followed by the Polar Vantage V (5.9 bpm), Garmin Fenix 5 (9.1 bpm), and Fitbit Versa (13.3 bpm).

For monitoring HR when running at 3.6 m/s or faster, performing intermittent sprints, or with a HR of 153-177 bpm, the Apple Watch Series 4 again exhibited the highest validity (average 6.0 bpm deviation from the criterion measure), followed by the Polar Vantage V (8.5 bpm), Fitbit Versa (8.8 bpm), and Garmin Fenix 5 (11.0 bpm).

Overall, when measuring HR, the Apple Watch Series 4 was the most valid (average 3.9 bpm deviation from the criterion measure), followed by the Polar Vantage V (7.0 bpm), Garmin Fenix 5 (9.9 bpm), and Fitbit Versa (11.4 bpm).

The validity of HR monitoring by the Apple Watch Series 4 and Polar Vantage V tended to be influenced less by the exercise intensity than that with the Garmin Fenix 5 and Fitbit Versa.

On average, all 4 wearables were poor at monitoring EE at the tested intensities and durations. The Apple Watch Series 4 deviated from the criterion measure by 124 kcal/h (CV 21%), Polar Vantage V by 121 kcal/h (CV 20%), Garmin Fenix 5 by 131 kcal/h (CV 22%), and Fitbit Versa by 112 kcal/h (CV 19%): average for the different intensities, with extrapolation of the CV for the 5-minute measurements to 1 hour.

To the best of our knowledge, this is the first assessment of the validity of these specific wrist-worn wearables. This is not surprising, since companies rarely rigorously validate new wearable models [20,21]. Comparison of our findings to earlier models requires caution, since it is not known whether the sensors or algorithms have been changed. However, such comparison might be of value to the manufacturers and to generally estimate if the parameters provided by the different manufacturers tend to be valid.

### Heart Rate Measurement

Previous comparison of earlier models of wrist-worn wearables sold by Apple, Polar, Garmin, and Fitbit at different intensities concluded that the Apple Watch Series 2 demonstrated the best validity for monitoring HR during exercise, followed by the Polar A380, Fitbit Blaze, Fitbit Charge 2, and Garmin Vivosmart HR, in that order, with absolute mean percentage errors of 4.1%, 19.5%, 21.1%, 21.4%, and 25.4%, respectively [32].

Another earlier comparison of the error rates of the Apple Watch (version not indicated), Fitbit Charge HR, and Garmin Forerunner 225 during light and vigorous running on a treadmill found that the Apple Watch displayed the highest validity (mean absolute percentage error of 1.1%-6.7%), followed by the Fitbit Charge HR (2.4%-17.0%) and Garmin Forerunner 225 (7.8%-24.4%) [33].

In addition, Thomson et al [34] validated HR measurements from the Fitbit Charge HR2 and Apple Watch of 30 young adults performing the Bruce Protocol and concluded that the relative error rates of the latter (2.4%-5.1%) were lower than for the

Fitbit wearable (3.9%-13.5%) at all the investigated exercise intensities.

Thus, these previous and our present findings indicate that the wrist-worn wearables made by Apple Inc and Polar Electro Oy exhibit the highest validity for measuring HR during physical activity at different levels, followed by Garmin or Fitbit wearables. However, additional comparative studies with different populations and different activities are required.

### Energy Expenditure

The majority of the sTEE values for the EE values provided by all the wearables were large, very large, or extremely large. Even though the Apple Watch Series 4 had the best validity, its sTEE values ranged from moderate to very large, while those for the Polar Vantage V, Garmin Fenix 5, and Fitbit Versa ranged from moderate to extremely large, with no apparent dependency on exercise intensity. Since these error rates exceed acceptable levels of validity, we cannot determine whether the unpredictable arm movements associated with the intermittent multidirectional sprint protocol affected the validity.

Thus, utilization of these wearables by researchers monitoring EE during interventions designed to increase physical activity is likely to lead to flawed conclusions. They would not assist with enhancing physical activity or counteracting noncommunicable diseases and would instead endanger the trustworthiness of applying consumer grade wearables to improve health.

These findings of the poor validity of wrist-worn wearables for monitoring EE are in line with previous reports. Bai et al [35] found that the Apple Watch Series 1 had a smaller mean absolute percentage error (15.2%) when assessing EE than the Fitbit Wearable (32.9%), both when sedentary and during aerobic and light-to-vigorous physical activity [35].

Wahl et al [25] concluded that none of the 11 wrist-worn wearables they investigated, including devices from Garmin and Fitbit, should be used to monitor EE while performing activities of intensities similar to those investigated here. In a systematic review published in 2015, Evenson et al [21] stated that the validity of wearables for monitoring EE is low.

At the same time, when Kinnunen et al [36] aimed to assess the long-term validity of wrist-worn motion sensors for monitoring daily EE, they were able to explain as much as 85% of the variation in total EE (compared to the double-labelled water procedure) by including HR during weekly exercises in their analysis. This indicates the potential usefulness of wrist-worn wearables for estimating EE.

In a previous study that took age, gender, body mass, and HR into account, the correlation coefficient for predicting EE during 10 minutes of exercise could be as high as 0.913 with a mixed model [37]. Considering the considerable validity of HR measurements by wearables and the ability to incorporate all the information required into an appropriate algorithm, we believe that more precise estimation of EE by the wearables examined here should be feasible.

However, our findings and most of the available scientific literature indicate that the wearables investigated here should not be employed to estimate EE at these exercise intensities for the durations assessed. Here, we monitored EE for <5 minutes, since countries such as the United States or Australia promote such short periods of physical activity in their guidelines [38,39]. In this context, certain studies have demonstrated positive effects of even very brief vigorous exercise, such as walking up a staircase 3 times on 3 separate days each week for 6 weeks [40]. Whether these devices can be used to monitor EE reliably over longer time periods remains to be determined.

Our experiment involved Caucasians performing light-to-vigorous exercise on a treadmill under laboratory conditions, and extrapolation of our findings to other populations or settings (eg, cycling, rowing, strength training) must be performed with caution [22]. For example, skin color may influence assessment of HR by photoplethysmography. Moreover, since our participants performed either light or vigorous physical activity, we cannot draw conclusions about validity at moderate levels.

We wish to emphasize that our current findings only apply to the specific modes of the wearables we used (eg, the “indoor walking mode” for the Apple Watch) selected for the different physical activities and that other modes might give different results. The Apple Watch Series 4 and Polar Vantage V allow selection of more differentiated modes of activity (eg, the “indoor walking” and “indoor running” modes were selected on the Apple Watch for the corresponding activities) than the Garmin Fenix 5 and Fitbit Versa (for which the “Treadmill” mode was selected for all activities).

## Conclusions

For measuring HR while sitting or during light-to-vigorous physical activity, the Apple Watch Series 4 exhibited the best validity (ie, the smallest error rates), followed by the Polar Vantage V, Garmin Fenix 5, and Fitbit Versa, in that order. The Apple Watch Series 4 and Polar Vantage V can be used for valid HR measurements at the intensities tested, whereas HR acquired with the Garmin Fenix 5 and Fitbit Versa must be interpreted cautiously due to their higher rates of error.

None of these wrist-worn wearables should be used to monitor EE at the intensities and durations tested.

## Acknowledgments

We would like to thank Engelhorn Sports GmbH for their support. This study was funded by the German Research Foundation (DFG) and the University of Wuerzburg in the funding program Open Access Publishing. This study was partly funded by the excellence initiative “Field of Focus 4: Self-Regulation and Regulation” at the Heidelberg University (Grant No. ZUK49/2 Ü).

## Conflicts of Interest

BS received funding from Polar Electro in connection with a previous project unrelated to the present investigation. The results of this study are presented clearly, honestly, and without fabrication, falsification, or inappropriate manipulation of data.

## References


1. World Health Organization. 2018 Feb 23. Physical activity URL: http://www.who.int/mediacentre/factsheets/fs385/en/ [accessed 2020-03-07]
2. World Health Organization. 2010. Global recommendations on physical activity for health URL: https://www.who.int/dietphysicalactivity/factsheet_recommendations/en/ [accessed 2020-03-07]
3. World Health Organization. World Health Organization. 2020 Mar 03. Obesity and overweight URL: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2020-03-07]
4. World Health Organization. 2018. Global action plan on physical activity 2018?2030: more active people for a healthier world URL: https://www.who.int/ncds/prevention/physical-activity/global-action-plan-2018-2030/en/ [accessed 2020-03-07]
5. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
6. Trost SG. Objective measurement of physical activity in youth: current issues, future directions. Exerc Sport Sci Rev 2001;29(1):32-36. [doi: 10.1097/00003677-200101000-00007] [Medline: 11210445]
7. Seiler S. What is best practice for training intensity and duration distribution in endurance athletes? Int J Sports Physiol Perform 2010 Sep;5(3):276-291. [doi: 10.1123/ijspp.5.3.276] [Medline: 20861519]
8. Swain DP, Abernathy KS, Smith CS, Lee SJ, Bunn SA. Target heart rates for the development of cardiorespiratory fitness. Medicine & Science in Sports & Exercise 1994;26(1):112-116. [doi: 10.1249/00005768-199401000-00019]
9. Donnelly JE, Blair SN, Jakicic JM, Manore MM, Rankin JW, Smith BK. Appropriate Physical Activity Intervention Strategies for Weight Loss and Prevention of Weight Regain for Adults. Medicine & Science in Sports & Exercise 2009;41(2):459-471. [doi: 10.1249/mss.0b013e3181949333]
10. KOEHLER K, BRAUN H, DE MARÉES M, FUSCH G, FUSCH C, SCHAENZER W. Assessing Energy Expenditure in Male Endurance Athletes. Medicine & Science in Sports & Exercise 2011;43(7):1328-1333. [doi: 10.1249/mss.0b013e31820750f5]
11. Waldrop MM. The chips are down for Moore's law. Nature 2016 Feb 11;530(7589):144-147. [doi: 10.1038/530144a] [Medline: 26863965]
12. Thompson WR. WORLDWIDE SURVEY OF FITNESS TRENDS FOR 2019. ACSM s Health & Fitness Journal 2018;22(6):10-17. [doi: 10.1249/fit.0000000000000438]
13. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
14. Düking P, Achtzehn S, Holmberg H, Sperlich B. Integrated Framework of Load Monitoring by a Combination of Smartphone Applications, Wearables and Point-of-Care Testing Provides Feedback that Allows Individual Responsive Adjustments to Activities of Daily Living. Sensors (Basel) 2018 May 19;18(5):1632 [FREE Full text] [doi: 10.3390/s18051632] [Medline: 29783763]
15. Techniker Krankenkasse. URL: https://www.tk.de/tk/themen/digitale-gesundheit/gesundheitsfoederung-durch-fitnesstracker-interview-dr-jens-baas/931248 [accessed 2018-04-04] [WebCite Cache ID 6yQE4yKX4]
16. World Health Organization. Addressing Mobile Health URL: https://www.who.int/ncds/prevention/be-healthy-be-mobile/en/ [accessed 2020-02-25]
17. Deutscher Bundestag. 2019 Sep 23. Entwurf eines Gesetzes für eine bessere Versorgung durch Digitalisierung und Innovation URL: https://www.bundesgesundheitsministerium.de/fileadmin/Dateien/3_Downloads/Gesetze_und_Verordnungen/GuV/D/DVG_Bundestag.pdf [accessed 2020-03-07]
18. Sperlich B, Holmberg H. Wearable, yes, but able…?: it is time for evidence-based marketing claims!. Br J Sports Med 2017 Aug 16;51(16):1240-1240 [FREE Full text] [doi: 10.1136/bjsports-2016-097295] [Medline: 27986762]
19. IDC. 2019 Sep 12. Wrist-Worn Wearables Maintain a Strong Growth Trajectory in Q2 2019, According to IDC URL: https://www.idc.com/getdoc.jsp?containerId=prUS45521319 [accessed 2019-11-29]
20. Düking P, Hotho A, Holmberg H, Fuss FK, Sperlich B. Comparison of Non-Invasive Individual Monitoring of the Training and Health of Athletes with Commercially Available Wearable Technologies. Front Physiol 2016 Mar 09;7:71 [FREE Full text] [doi: 10.3389/fphys.2016.00071] [Medline: 27014077]
21. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]

22. Düking P, Fuss FK, Holmberg H, Sperlich B. Recommendations for Assessment of the Reliability, Sensitivity, and Validity of Data Provided by Wearable Sensors Designed for Monitoring Physical Activity. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e102 [FREE Full text] [doi: 10.2196/mhealth.9341] [Medline: 29712629]
23. Düking P, Born DP, Sperlich B. The SpeedCourt: Reliability, Usefulness, and Validity of a New Method to Determine Change-of-Direction Speed. Int J Sports Physiol Perform 2016 Jan;11(1):130-134. [doi: 10.1123/ijspp.2015-0174] [Medline: 26061971]
24. Vogler AJ, Rice AJ, Gore CJ. Validity and reliability of the Cortex MetaMax3B portable metabolic system. J Sports Sci 2010 May;28(7):733-742. [doi: 10.1080/02640410903582776] [Medline: 20419553]
25. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-Validity of Commercially Available Physical Activity Tracker to Estimate Step Count, Covered Distance and Energy Expenditure during Sports Conditions. Front Physiol 2017 Sep 22;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
26. Hermand E, Cassirame J, Ennequin G, Hue O. Validation of a Photoplethysmographic Heart Rate Monitor: Polar OH1. Int J Sports Med 2019 Jul 12;40(7):462-467. [doi: 10.1055/a-0875-4033] [Medline: 31189190]
27. Schubert M, Clark A, De La Rosa A. The Polar OH1 Optical Heart Rate Sensor is Valid during Moderate-Vigorous Exercise. Sports Med Int Open 2018 Jun 17;2(3):E67-E70 [FREE Full text] [doi: 10.1055/a-0631-0920] [Medline: 30539120]
28. Hopkins W. Spreadsheets for analysis of validity and reliability. Sportscience 2017;21.
29. Khushhal A, Nichols S, Evans W, Gleadall-Siddall D, Page R, O'Doherty A, et al. Validity and Reliability of the Apple Watch for Measuring Heart Rate During Exercise. Sports Med Int Open 2017 Oct 18;1(6):E206-E211 [FREE Full text] [doi: 10.1055/s-0043-120195] [Medline: 30539109]
30. Hopkins W. Validity thresholds and error rates for test measures used to assess individuals. In: Proceedings of the 21st Annual Congress of the European College of Sport Science. 2016 Jun 6 Presented at: 21st Annual Congress of the European College of Sport Science; 2016; Vienna, Austria.
31. Ainsworth BE, Haskell WL, Whitt MC, Irwin ML, Swartz AM, Strath SJ, et al. Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 2000 Sep;32(9 Suppl):S498-S504. [doi: 10.1097/00005768-200009001-00009] [Medline: 10993420]
32. BOUDREAUX BD, HEBERT EP, HOLLANDER DB, WILLIAMS BM, CORMIER CL, NAQUIN MR, et al. Validity of Wearable Activity Monitors during Cycling and Resistance Exercise. Medicine & Science in Sports & Exercise 2018;50(3):624-633. [doi: 10.1249/mss.0000000000001471]
33. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
34. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019 Jun 18;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
35. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug 06;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
36. Kinnunen H, Häkkinen K, Schumann M, Karavirta L, Westerterp KR, Kyröläinen H. Training-induced changes in daily energy expenditure: Methodological evaluation using wrist-worn accelerometer, heart rate monitor, and doubly labeled water technique. PLoS One 2019 Jul 10;14(7):e0219563 [FREE Full text] [doi: 10.1371/journal.pone.0219563] [Medline: 31291373]
37. Keytel L, Goedecke J, Noakes T, Hiiloskorpi H, Laukkanen R, van der Merwe L, et al. Prediction of energy expenditure from heart rate monitoring during submaximal exercise. J Sports Sci 2005 Mar 18;23(3):289-297. [doi: 10.1080/02640410470001730089] [Medline: 15966347]
38. The Department of Health, Australian Government. 2019 Apr 12. Australia's Physical Activity and Sedentary Behaviour Guidelines and the Australian 24-Hour Movement Guidelines URL: https://www1.health.gov.au/internet/main/publishing.nsf/Content/health-pubhlth-strateg-phys-act-guidelines [accessed 2020-07-03]
39. US Department of Health and Human Services. 2018. Physical Activity Guidelines for Americans 2nd Edition URL: https://health.gov/paguidelines/second-edition/pdf/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2020-03-07]
40. Jenkins EM, Nairn LN, Skelly LE, Little JP, Gibala MJ. Do stair climbing exercise "snacks" improve cardiorespiratory fitness? Appl Physiol Nutr Metab 2019 Jun;44(6):681-684. [doi: 10.1139/apnm-2018-0675] [Medline: 30649897]


## Abbreviations

**CV:** coefficient of variation.
**EE:** energy expenditure.
**HR:** heart rate.
**MET:** metabolic equivalent.
**sTEE:** standardized typical error of the estimate.
**WHO:** World Health Organization.

*Edited by G Eysenbach; submitted 21.10.19; peer-reviewed by L Ardigò, E Kemler; comments to author 25.11.19; revised version received 12.12.19; accepted 24.01.20; published 06.05.20.*

*Please cite as:*
*Düking P, Giessing L, Frenkel MO, Koehler K, Holmberg HC, Sperlich B*
*Wrist-Worn Wearables for Monitoring Heart Rate and Energy Expenditure While Sitting or Performing Light-to-Vigorous Physical Activity: Validation Study*
*JMIR Mhealth Uhealth 2020;8(5):e16716*
*URL: https://mhealth.jmir.org/2020/5/e16716*
*doi:10.2196/16716*
*PMID:32374274*

Original Paper

# Low-Cost Consumer-Based Trackers to Measure Physical Activity and Sleep Duration Among Adults in Free-Living Conditions: Validation Study

Laurent Degroote[1,2,3], MSc; Gilles Hamerlinck[1], BSc; Karolien Poels[4], PhD; Carol Maher[5], PhD; Geert Crombez[2], PhD; Ilse De Bourdeaudhuij[1], PhD; Ann Vandendriessche[6], MSc; Rachel G Curtis[5], PhD; Ann DeSmet[1,3,4,7], PhD

[1]Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium

[2]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

[3]Research Foundation Flanders, Brussels, Belgium

[4]Department of Communication Studies, University of Antwerp, Antwerp, Belgium

[5]School of Health Sciences, University of South Australia, Adelaide, Australia

[6]Department of Public Health and Primary Care, Ghent University, Ghent, Belgium

[7]Department of Clinical and Health Psychology, Université Libre de Bruxelles, Brussels, Belgium

**Corresponding Author:**
Laurent Degroote, MSc
Department of Movement and Sports Sciences
Ghent University
Watersportlaan 2
Ghent
Belgium
Phone: 32 9 264 62 99
Email: laurent.degroote@ugent.be

## Abstract

**Background:** Wearable trackers for monitoring physical activity (PA) and total sleep time (TST) are increasingly popular. These devices are used not only by consumers to monitor their behavior but also by researchers to track the behavior of large samples and by health professionals to implement interventions aimed at health promotion and to remotely monitor patients. However, high costs and accuracy concerns may be barriers to widespread adoption.

**Objective:** This study aimed to investigate the concurrent validity of 6 low-cost activity trackers for measuring steps, moderate-to-vigorous physical activity (MVPA), and TST: Geonaut On Coach, iWown i5 Plus, MyKronoz ZeFit4, Nokia GO, VeryFit 2.0, and Xiaomi MiBand 2.

**Methods:** A free-living protocol was used in which 20 adults engaged in their usual daily activities and sleep. For 3 days and 3 nights, they simultaneously wore a low-cost tracker and a high-cost tracker (Fitbit Charge HR) on the nondominant wrist. Participants wore an ActiGraph GT3X+ accelerometer on the hip at daytime and a BodyMedia SenseWear device on the nondominant upper arm at nighttime. Validity was assessed by comparing each tracker with the ActiGraph GT3X+ and BodyMedia SenseWear using mean absolute percentage error scores, correlations, and Bland-Altman plots in IBM SPSS 24.0.

**Results:** Large variations were shown between trackers. Low-cost trackers showed moderate-to-strong correlations (Spearman $r$=0.53-0.91) and low-to-good agreement (intraclass correlation coefficient [ICC]=0.51-0.90) for measuring steps. Weak-to-moderate correlations (Spearman $r$=0.24-0.56) and low agreement (ICC=0.18-0.56) were shown for measuring MVPA. For measuring TST, the low-cost trackers showed weak-to-strong correlations (Spearman $r$=0.04-0.73) and low agreement (ICC=0.05-0.52). The Bland-Altman plot revealed a variation between overcounting and undercounting for measuring steps, MVPA, and TST, depending on the used low-cost tracker. None of the trackers, including Fitbit (a high-cost tracker), showed high validity to measure MVPA.

**Conclusions:** This study was the first to examine the concurrent validity of low-cost trackers. Validity was strongest for the measurement of steps; there was evidence of validity for measurement of sleep in some trackers, and validity for measurement of MVPA time was weak throughout all devices. Validity ranged between devices, with Xiaomi having the highest validity for

measurement of steps and VeryFit performing relatively strong across both sleep and steps domains. Low-cost trackers hold promise for monitoring and measurement of movement and sleep behaviors, both for consumers and researchers.

***(JMIR Mhealth Uhealth 2020;8(5):e16674)*** doi:10.2196/16674

**KEYWORDS**

fitness trackers; mobile phone; accelerometry; physical activity; sleep

## Introduction

### Background

Physical activity (PA) and sleep are modifiable determinants of morbidity and mortality among adults and specifically contribute to the development of diseases such as obesity, type 2 diabetes, cardiovascular diseases, low quality of life, and mental health problems [1-6]. Engaging in at least 30 min of moderate-to-vigorous physical activity (MVPA) per day, getting between 7 and 9 hours of total sleep time (TST) per night, and spending relatively more time on light PA rather than being sedentary are associated with beneficial health outcomes [1-6]. A large proportion of adults do not meet the guidelines for one or more of these behaviors [6,7]. PA and sleep are, together with time spent on sedentary behavior (SB), codependent behaviors: they are part of one 24-hour day, and time spent on one behavior will impact the time spent on at least one of the other behaviors. It is, therefore, recommended to target these behaviors together [8].

Successful health promotion interventions rely on behavior change techniques that address modifiable determinants of health behavior [9]. A behavior change technique reported as both effective [10] and highly appreciated by users [11,12] is self-monitoring of health behavior. Self-monitoring refers to keeping a record of the behavior that is performed [13]. Self-monitoring tools provide opportunities for self-management of health as well as for remote activity tracking by health care providers as part of a patient's treatment regimen [14]. Subjective ways of self-monitoring, such as self-report using retrospective measures (eg, diaries and questionnaires), often come with high participant burden and reporting biases [14]. Self-reported sleep duration in sleep logs showed an overestimation in comparison with objective measurements, especially when sleep duration was below the recommended health norms [15]. Activity trackers conversely offer automated, objective, and convenient means for self-monitoring PA and sleep. This paper focused on self-monitoring via consumer-based activity trackers as intervention tools for PA and sleep, more specifically by investigating the validity of low-cost trackers. Such trackers rarely monitor SB [16,17], which is why SB, although important in 24-hour movement behaviors, falls outside the scope of this paper.

Activity trackers may include pedometers, smartphone-based accelerometers, and accelerometers in advanced electronic wearable trackers or in smartwatches. However, pedometers do not provide information on sleep, and smartphone-based accelerometers have shown lower accuracy when measuring PA compared to advanced electronic wearable trackers [18], making advanced electronic wearable trackers and smartwatches more suitable to accurately self-monitor PA and sleep. Smartwatches (eg, Apple Watch) offer several other functions apart from activity tracking such as communication and entertainment and are usually more expensive than advanced electronic wearable trackers (eg, Fitbit Charge). Advanced electronic wearable trackers (termed as activity trackers hereafter) are usually wrist-or belt-worn, provide 24-hour self-monitoring, and often include real-time behavioral feedback or more detailed feedback shown after synchronization with other electronic devices (eg, tablet, smartphone, or PC) [19]. Several commercial activity trackers are available to the public and are increasingly integrated into effective intervention programs to improve activity behaviors [20,21].

There has been an increased interest by adults in activity trackers. For example, in Flanders, Belgium, 8% of adults owned an activity tracker (22% owned a type of wearable, including sports watches and smartwatches) in 2018 compared with only 2% (8% owned a type of wearable, including sports watches and smartwatches) owning one in 2015 [22]. Characteristics of activity trackers may impact their continued use and further adoption. Cost is likely to be a barrier to increased adoption of higher-end trackers [19,23]. Indeed, activity trackers appear to be used less among adults who are less educated, unemployed [19], and have a lower income [22]. Notably, unhealthy lifestyles such as insufficient PA [24] and insufficient sleep duration [25] are more prevalent among people of lower to medium socioeconomic status (SES) than among those of higher SES. Therefore, providing accurate, low-cost options to self-monitor PA and sleep in their daily lives is crucial for public health, as a lack of valid low-cost trackers may increase the health and digital divide between lower and higher income groups in the society. However, nonadoption of activity trackers in low SES populations can probably not only be attributed to the high cost of the devices but may also be a matter of priorities and affordances. Further research in this area is necessary. Having valid low-cost trackers not only plays a role in low SES populations but also in the general population; cost-effective solutions are needed for scaling up interventions in a public health context where financial resources are limited [26]. Having accurate, low-cost activity trackers can be expected to increase the feasibility of scaling up interventions that rely on activity trackers.

The unequal access to valid tools because of cost barriers is often studied within health literacy conceptual frameworks. Health literacy refers to having the ability and motivation to take responsibility for one's own health [27]. Low health literacy has been associated with worse health outcomes [27], and improving access to tools that can help understand their own health behavior via self-monitoring and taking responsibility to take care of one's own health may improve health literacy. There is increasing attention to expanding the health literacy

model to electronic health (eHealth) literacy or digital health literacy, defined as the ability of people to use emerging technology tools to improve or enable health and health care [28]. Digital health literacy appears to be associated with a lower SES [29].

More specifically, the importance of accuracy of low-cost trackers can also be understood from the technology acceptance model that emphasizes the need for trust and perceived usefulness, together with perceived ease of use, of a tool before users are willing to adopt them [30].

When using activity trackers, their accuracy needs to be established to avoid counterproductive effects, such as falsely signaling that people are meeting guidelines and need not make any extra efforts, whereas, in fact, these people may not reach the sufficient sleep or PA levels [31]. Conversely, an underestimation of actual behavior can also cause people to get demotivated and to no longer make efforts to do better [32]. Accuracy of the tracker has also been cited by users as the trackers' most important characteristic [19]. To effectively use wearable activity trackers for health self-management in daily life, accuracy needs to be assessed in free-living settings because laboratory-based validity studies tend to overestimate validity [18]. The validity to measure PA in free-living conditions has been examined for several activity trackers, such as Fitbit One, Zip, Ultra, Classic, Flex [16-18,33-38], Misfit Shine [18], and Withings Pulse [18]. In general, studies found the highest validity for Fitbit trackers [18]. Most validated trackers showed high correlations with an ActiGraph accelerometer for number of steps [16,39,40]. MVPA is less often studied and less accurately measured by activity trackers than step count [39]. Activity trackers showed moderate-to-strong correlations with ActiGraph accelerometers on MVPA, with Fitbit trackers and Withings Pulse showing the highest accuracy [39]. In addition, for TST, several wearable activity trackers currently on the market have been assessed for validity, including Fitbit (Flex and Charge HR) [41-43], Withings Pulse [39,41], Basis Health Tracker [41], Garmin [44], and Polar Loop [44]. Validity results for TST were very divergent, ranging from low to strong validity, with Fitbit again showing better validity [39,44]. The accuracy of PA and/or TST depends on the position where the tracker is worn, for example, the wrist vs the hip [33], and can be improved by combining accelerometry with the heart rate measurement [45,46].

The cost of the trackers in the abovementioned published validation studies was often not reported, but their price in the current market (at the end of January 2019) ranged from €50 (US $56) to €130 (US $146) for an unused, basic model (with Misfit Flash as the exception at €42 [US $47]). Most trackers that are popular in the consumer market and that are reported on in scientific publications cost more than €50 (US $56) and commonly more than €100 (US $112) [14]. A recent industry report states that when spending less than US $50, users are likely to get a product of mediocre accuracy [47], although it is unclear whether this statement was empirically based. To our knowledge, only 2 studies have examined the validity of low-cost trackers. Wahl et al's [48] study of the Polar Loop (price in June 2019 around €60 [US $67]), Beurer AS80 (price in June 2019 around €42 [US $47]), and Xiaomi Mi Band (price in June 2019 around €25 [US $28]) suggested that only the Mi Band had good validity for step count. However, this study was conducted in a laboratory and not in free-living conditions. In one other study, the validity of the Xiaomi Mi Band for measuring TST was evaluated relative to a manual switch-to-sleep-mode measurement, with positive results [49]. However, this study did not use an objective measurement tool for comparison. We are not aware of any validation studies of low-cost activity trackers against objective measurement methods conducted in free-living conditions, and many of the most commonly available low-cost trackers do not appear to have been validated in any form.

In summary, wearable activity trackers can be a useful tool in health promotion and remote treatment monitoring for PA and TST. However, high costs and accuracy concerns may be barriers to widespread adoption [50]. Assessing the validity of low-cost trackers may play a major role at the population level to encourage health behavior in the future and among low SES groups who are most at risk for poor health and in need of healthy behavior promotion. To enable activity self-monitoring in daily life, the accuracy of low-cost wearable activity trackers needs to be established in free-living conditions. Current validation studies have mainly focused on wearable activity trackers that cost above €50 (US $56).

### Objectives

This study aimed to assess the validity of low-cost wearable activity trackers among adults (≤€50 [US $56]) for the objective measurement of PA and TST in daily life against free-living gold standards (ActiGraph GT3X+ accelerometer and BodyMedia SenseWear). This study was exploratory in nature and did not have firm hypotheses regarding the validity of specific low-cost trackers. However, it may be expected that trackers with heart rate monitoring are more accurate than those without heart rate monitoring. This may be because heart rate measurement contributes to a more accurate estimate of intensity and energy expenditure, resulting in a more accurate discrimination between activity and nonactivity [45,46].

## Methods

### Participants and Procedure

A concurrent validity study among adults was designed in which a low-cost tracker was validated against a free-living condition standard for steps, active minutes (MVPA), and TST. A high-cost tracker (Fitbit Charge 2) was also validated against these gold standards, to compare with validation outcomes for the low-cost trackers. In each participant, three 24-hour observation days were collected for each low-cost tracker. Power analyses (run in G*Power 3.1.9.2) suggested that to detect a 2-tailed significant correlation ($H_1$) of 0.49 to 0.90, with 80% power (values based on the study by Brooke et al [44]), a sample size of between 6 and 29 was required.

A total of 20 healthy participants aged between 18 and 65 years living in Flanders, Belgium, were recruited using convenience sampling. Inclusion criteria were having no current physical limitations, medical conditions, or psychiatric conditions that may impact movement or sleep. Descriptive information

collected on participants consisted of age, sex, self-reported height and weight, and highest attained education. All participants read and signed an informed consent form. The Ethics Committee of the University Hospital of Ghent approved the study protocol (B670201731732).

## Instruments

### *Convergent Measure*

As this is a free-living study, the ActiGraph GT3X+ (ActiGraph) triaxial accelerometer was used as a reliable and valid reference for measuring step count [51-53] and MVPA [54,55]. The GT3X+ has been shown to be a valid measure of both step count compared with direct observation (percentage error <1.5% [52]; percentage error ≤1.1% [53]; and intraclass correlation coefficient [ICC]≥0.84 [51]) and MVPA compared with indirect calorimetry ($r$=0.88) [54]. Accelerometer data were initialized, downloaded, and processed using ActiLife version 5.5.5 software (ActiGraph). The Freedson Adult cut-points were applied to categorize PA measured by using the ActiGraph accelerometer (sedentary activity=0-99 counts per min, light activity=100-1951 counts per min, moderate activity=1952-5723 counts per min, and vigorous activity ≥5724 counts per min) [54]. A 15-second epoch was used when downloading the data. The ActiGraph GT3X+ was fitted to the right side of the participants' waist in accordance with the manufacturer's instructions. Only days with valid data of the ActiGraph were included in the analysis. A valid day was defined as a 24-hour period in which at least 10 hours of data wear time were recorded [56]. Nonwear time was analyzed as a run of zero counts lasting more than 60 min with an allowance of 2 min of interruptions. Using this algorithm, the risk of misclassification of nonwear time as sedentary time was avoided [57].

The BodyMedia SenseWear (BodyMedia Inc) is a portable multisensor device that can provide information regarding the total energy expenditure, TST, circadian rhythm, and other activity metrics. In this study, the SenseWear was used as the reference for sleep duration. SenseWear has been validated as a measure of TST compared with polysomnography ($r$=0.83; SE of estimate 37.71) [58]. Data were analyzed in SenseWear Professional 8.1 software [59]. The SenseWear was placed over the triceps muscle on the nondominant arm between the acromion and olecranon processes, in accordance with the manufacturer's instructions.

### *Low-Cost Activity Trackers*

In total, 6 low-cost activity trackers were selected (Figure 1) based on their price at the time of the study (≤€50 [US $56]), their market share (eg, MyKronoz and Xiaomi), whether or not they included a heart rate measurement and output (steps, MVPA or active minutes, and TST), and availability from popular web-based purchase sites in Europe where the study was conducted. Furthermore, we tested the Fitbit Charge 2 to also include a comparison between a low-cost activity tracker and a validated high-cost activity tracker. Fitbit was selected as a high-cost activity tracker because it was one of the most popular activity trackers on the market at the start of the study and was already validated for measuring steps, MVPA, and TST [17]. All participants received a Wiko smartphone in loan (Lenny 3, Android 6.0 Marshmallow, price €99.99 [US $119.80] in June 2019) to pair the trackers with, to cancel out any potential individual differences in smartphone pairing.

All devices measured steps and TST. Only Xiaomi, Nokia, and also Fitbit used a specific variable that quantifies intensive forms of PA. These 3 devices reported *active minutes* with no further subdivision. As all the devices set a goal of 30 min PA per day (similar to the MVPA recommendations for adults), it was assumed that the measured variable corresponded to MVPA as measured by the ActiGraph. However, specific information regarding intensity cut-points is not publicly available. TST was used, excluding daytime naps, for comparison with the SenseWear that was only worn at night. Only Fitbit, VeryFit, and Xiaomi measured the heart rate. Data were extracted using the proprietary software for all devices, in the same fashion that a consumer would use the software, and were visually checked for outliers.

**Figure 1.** Tracker characteristics.

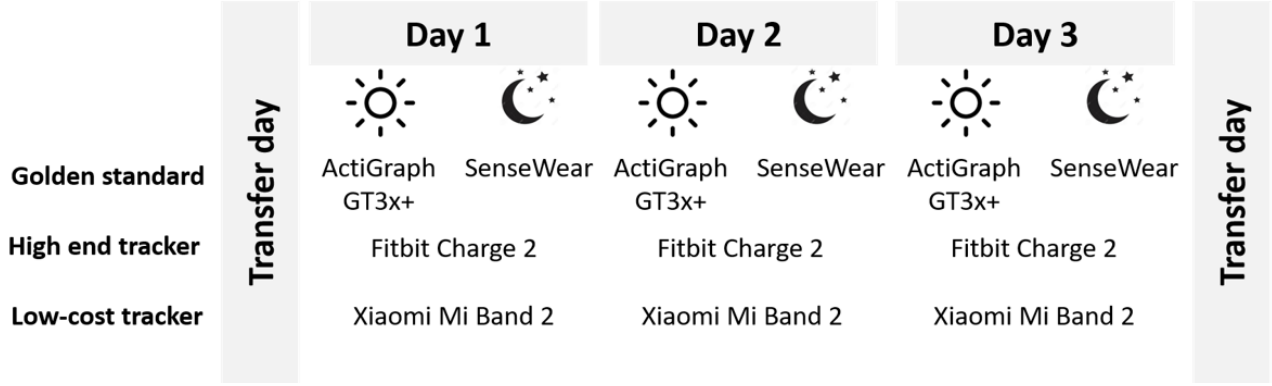
| | Low-cost trackers | | | | | | High-cost tracker |
|---|---|---|---|---|---|---|---|
| | **Geonaut OnCoach** | **iWowni5** | **MyKronoz ZeFit4** | **Nokia GO (previous-ly Withings GO)** | **VeryFit 2.0, ID 115** | **Xiaomi Mi Band2** | **Fitbit Charge 2** |
| **Image** |  |  |  |  |  |  |  |
| **Manufacturer** | Decathlon (°1976, Villeneuve-d'Ascq, France) | iWownFit (°2012, Shenzhen, China) | MyKronoz (°2013, Genève, Switzerland) | Nokia (°1978, Espoo, Finland) | Zencro Industrial Co.,Limited (°2017, Dongguan City, China) | Xiaomi Corporation (°2010, Peking, China) | Fitbit (°2007, San Francisco, California, USA) |
| **Global market share wearables (2018-2019)** | No market share info available | 1.5% (own website | No precise %;listed as #4 | No market share info available | No market share info available | 13.0% Q1 2019 | 6.0% Q1 2019;listed as #6 |
| **Rechargeable** | | ✓ | ✓ | | ✓ | ✓ | ✓ |
| **Price** | | | | | | | |
| at start study June 2019 | €24.99 (US$27.77) | €25.50 (US$28.34) | €48.00 (US$53.34) | €50.00 (US$55.56) | €29.95 (US$33.28) | €39.95 (US$44.40) | €125.00 (US$138.91) |
| **Output** | | | | | | | |
| Steps | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Active minutes (MVPA) | × | × | × | ✓ | × | ✓ | ✓ |
| Total sleep time | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Heart-rate** | × | × | × | × | ✓ | ✓ | ✓ |

## Free-Living Protocol

As it was not feasible or comfortable to wear all trackers at the same time, participants were instructed to wear one of the low-cost devices in combination with the Fitbit tracker on their nondominant wrist. They were also instructed to simultaneously wear the ActiGraph on their hip during daytime and the SenseWear on their upper arm at nighttime. Furthermore, the participants were provided with a diary to write down the time they put on and took off the devices. This way it could be checked that the devices were always worn simultaneously. If this was not the case, data of the device that was worn separately were deleted to avoid a mismatch of the measurements. Participants received the 6 low-cost trackers in a random order. The position of the low-cost and high-cost tracker on the nondominant wrist (first or second in distance from the wrist) was varied across days. Each tracker was worn for a period of 3 consecutive days and nights. A period of 3 days and 3 nights was chosen to balance between achieving sufficient data for the question under study without burdening the participants. Between 2 periods, a 1-day gap allowed for switching the devices. During daytime, the devices were worn during all waking hours, except during water-based activities. When participants went to bed, they were asked to remove the ActiGraph and put on the SenseWear instead. In Figure 2, a typical measurement period for one device is shown.

PA or TST may differ between weekdays and weekend days. Although this study did not intend to explain differences in PA or TST but rather the degree of agreement between 2 measurements on any given day, a difference in how often a tracker was measured on a certain day rather than another day may influence validity results. For example, validity has shown to be lower for measuring a low number of steps or high number of steps. Our study design controlled for this potential influence by randomly varying the days across participants on which a particular tracker was worn. Across all data points, we would then expect all measurement days to be relatively equally represented, as was the case in our study. The percentage of weekend days in total measurement days ranged between 25% and 33%. In addition, on particular weekdays, there were very few differences (2%-9% difference between the tracker with the lowest number of measurements on a certain day and the tracker with the highest number of measurements on a particular day).

**Figure 2.** An example of the measurement protocol for one period.




### Statistical Analysis

Analyses were performed using IBM SPSS Statistics version 24.0 (SPSS Inc). All analyses were performed on a daily measurement level, counting a measured day as a unit of analysis. Analyses consisted of measures of agreement, systematic differences, and bias and limits of agreement. Measures of agreement (equivalence testing) included the Spearman correlation coefficient (*r*) to examine the association between steps, active minutes, and TST measured by trackers and convergent measure (also illustrated in scatter plots). As sleep and PA data were nonnormally distributed, (1) a Spearman correlation, a nonparametric statistical test, was used instead of a Pearson correlation and (2) *ICC* (absolute agreement, 2-way random, single measures, 95% CI) that reflects the effect of individual differences on observed measures. Measures of systematic differences included mean absolute percentage errors (MAPEs) of tracker measurements compared with those of the convergent measure. MAPEs were calculated with the following formula: mean difference activity tracker−convergent measure × 100/mean gold standard. Bland-Altman plots with their associated limits of agreement were used to examine biases between measurements from the trackers and the convergent measure. The following cutoff values were used to interpret the Spearman correlation coefficient: <0.20=very weak, 0.20 to 0.39=weak, 0.40 to 0.59=moderate, 0.60 to 0.79=strong, and 0.80 to 1.0=very strong [60]. The cutoff values to interpret the ICC were as follows: <0.60=low, 0.60 to 0.75=moderate, 0.75 to 0.90=good, and >0.90=excellent [48].

A series of linear mixed effects models with restricted maximum likelihood estimation were used to examine the association between steps, MVPA minutes, and TST measured by the commercial trackers and convergent measures, accounting for the structure of the data (repeated measures clustered within participants). The pattern of results was similar to that obtained from the abovementioned analyses. Data are, therefore, presented in Multimedia Appendix 1.

## Results

### Descriptive Statistics

A total of 3 participants discontinued their participation in the study: one participant dropped out at the start of the study because of the combination of high perceived burden of the research protocol and a busy personal schedule, and consequently, no data were collected and analyzed from this participant; one participant was not able to meet the protocol toward the end of the study because of conflict with his/her work schedule; and one participant had to end participation because of an unexpected hospital admission (17/20, 85% retention rate). The average age of the analyzed sample of participants who started the study (n=19) was 37.6 (SD 13.4) years; of 19 participants, 13 were female. The sample was highly educated, with 17 participants having achieved a higher education degree (academic or nonacademic). Their average BMI was 23.5 (SD 4.4) kg/m$^2$. Two participants were overweight (BMI of 25-30 kg/m$^2$), and 2 participants were obese (BMI ≥30 kg/m$^2$). The level of MVPA measured at baseline with the International Physical Activity Questionnaire varied from 10 to 351 min per day (SD 91) [61,62].

All participants owned a smartphone; 5 of 19 participants had previous experience with wearable trackers (n=3; Fitbit). As can be expected in a highly educated sample, they were all very familiar with digital tools and required little assistance in installation or usage. We did not expect any impact of participants' experience on the validity measurements, as (1) these would not have a differential effect of any potential misuse between different trackers and (2) control procedures were put in place to prevent any misuse. Potential misuse could consist of a wrong placement of the tracker. Participants received a thorough briefing at the start of the study and a daily check-up of any issues to ensure any baseline differences in familiarity with digital tools were canceled out and to reduce the risk for misuse. No issues with misuse were noted.

### Issue of Usability With Low-Cost Trackers

In total, each device was intended to be tested for 60 days. As one of the participants did not start, the maximum number of potential measurement days per tracker was reduced to 57. The number of days of available data varied per tracker because of dropouts at the end of the study by some participants and because of technical issues experienced with some trackers, which resulted in fewer days of available data.

Of 57 measurement days, VeryFit had 55 (96%) measured days for PA (lost days: 2 because of no data shown in the app) and 51 (86%) measured days for sleep (lost days: 3 because of participant noncompliance and 3 because of no data shown in the app). Of 57 measurement days, iWown had 52 (89%) measured days for PA (lost days: 4 because the tracker did not

pair and 1 because of no data shown in the app) and 51 (89%) measured days for sleep (lost days: 4 because the tracker did not pair and 2 because of no data shown in the app). Xiaomi was not worn by 2 participants because of dropping out, reducing potential measurement days to 51. Of 51 measured days, Xiaomi had 48 (94%) measured days for PA (lost days: 2 because of participant noncompliance and 1 because of no data shown in the app) and 44 (86%) measured days for sleep (lost days: 6 because of participant noncompliance and 1 because of no data shown in the app). Of 57 measured days, Nokia had 49 (86%) measured days for PA (8 lost days because of no data shown in the app) and 46 (81%) measured days for sleep (lost days: 8 because of no data shown in the app and 3 because of participant noncompliance). MyKronoz was not worn by 3 participants because of dropping out; one participant accidentally removed the data, reducing potential measurement days to 45. Of 45 measured days, MyKronoz had 40 (89%) measurement days for PA (5 lost days because of no data shown in the app) and 24 (53%) for sleep (lost days: 11 because of no data shown in the app and 10 because of participant noncompliance). Of 57 measured days, Geonaut had 37 (65%) measured days for PA (lost days: 12 because of no data shown in the app, 9 because of the device not pairing, and 5 because of participant noncompliance) and 30 (53%) measured days for sleep (lost days: 9 days because of the device not pairing, 8 because of no data shown in the app, and 4 because of participant noncompliance).

Participants were especially frustrated about a device not pairing, as this meant they had to reinstall the tracker and also lost their past activity history. Thus, VeryFit and Xiaomi showed little data loss because of usability problems, whereas especially for Geonaut and MyKronoz, data were lost because of usability problems. In general, more data were lost for sleep than for PA. Usable data in the analyses were further reduced because of technical issues experienced with the convergent measures, which resulted in fewer days of data for which comparisons could be made (the number of usable data points are shown in all tables).

## Validity of Low-Cost Trackers

### *Physical Activity*

Table 1 shows the mean steps, mean minutes of MVPA, and the corresponding standard deviations for all trackers for measuring steps and MVPA.

**Table 1.** Mean steps and minutes of moderate-to-vigorous physical activity per day measured by the low-cost trackers, Fitbit and ActiGraph.

| Tracker | Number of measured days | Mean (SD) | Range |
|---|---|---|---|
| **Number of steps per day** | | | |
| Geonaut | 37 | 8026 (4352) | 657-19,413 |
| iWown | 51 | 7668 (5169) | 259-22,759 |
| MyKronoz | 40 | 10,431 (4764) | 485-24,493 |
| Nokia | 50 | 5896 (3113) | 325-13,976 |
| VeryFit | 55 | 7320 (4481) | 649-22,628 |
| Xiaomi | 48 | 7317 (4535) | 369-20,866 |
| Fitbit | 307 | 9662 (4866) | 451-24,664 |
| ActiGraph | 316 | 8126 (4314) | 188-23,121 |
| **Number of minutes of moderate-to-vigorous physical activity per day** | | | |
| Nokia | 49 | 5 (12) | 0-52 |
| Xiaomi | 46 | 80 (48) | 0-190 |
| Fitbit | 305 | 45 (49) | 0-239 |
| ActiGraph | 328 | 41 (31) | 0-150 |

Agreement testing for steps diverged between the Spearman *r* coefficient and ICC (Table 2). All trackers, except iWown, showed strong (Nokia, Geonaut, VeryFit, and MyKronoz) to very strong (Xiaomi and Fitbit) agreement with the ActiGraph measurements based on the Spearman *r* coefficient (all above 0.60). On the basis of ICC, MyKronoz, iWown, and Nokia showed low agreement (ICC<0.60), whereas Geonaut had moderate and Xiaomi, Fitbit, and VeryFit had a good agreement with the ActiGraph measurements (ICC=0.75-0.90). These coefficients are in line with the interpretation of the MAPE scores, showing the largest mean deviation from the ActiGraph measurements for iWown (35.28%) and the smallest for the Xiaomi tracker (17.14%).

For measuring MVPA, correlations between the MVPA measurements of the trackers and the ActiGraph accelerometer were weak for Nokia and Xiaomi and moderate for Fitbit (Table 2). The ICC showed low agreement for MVPA between all 3 trackers and the ActiGraph accelerometer (ICC<0.60). The MAPE scores also indicate very large mean deviations from the ActiGraph measurements for MVPA (>100%), which confirm the low accuracy of the trackers for measuring MVPA.

**Table 2.** Correlation coefficients, intraclass correlation coefficients, associated 95% CI of the measurements, and mean absolute percentage error scores for measuring steps and moderate-to-vigorous physical activity.

| Tracker | n | Spearman *r* (95% CI) | Intraclass correlation coefficient | 95% CI | Mean absolute percentage error (%) |
|---|---|---|---|---|---|
| **Steps** | | | | | |
| Geonaut | 36 | 0.63[a] (0.31 to 0.87) | 0.68[a] | 0.46 to 0.82 | 24.63 |
| iWown | 50 | 0.53[a] (0.16 to 0.77) | 0.51[a] | 0.28 to 0.69 | 35.28 |
| MyKronoz | 38 | 0.77[a] (0.45 to 0.95) | 0.59[a] | 0.22 to 0.79 | 25.79 |
| Nokia | 50 | 0.77[a] (0.51 to 0.94) | 0.56[a] | 0.27 to 0.74 | 22.62 |
| VeryFit | 54 | 0.78[a] (0.61 to 0.89) | 0.82[a] | 0.62 to 0.91 | 24.87 |
| Xiaomi | 45 | 0.91[a] (0.81 to 0.97) | 0.90[a] | 0.77 to 0.95 | 17.14 |
| Fitbit | 300 | 0.91[a] (0.86 to 0.94) | 0.87[a] | 0.66 to 0.93 | 25.73 |
| **Moderate-to-vigorous physical activity** | | | | | |
| Nokia | 16 | 0.24 (−0.11 to 0.50) | 0.18 | −0.10 to 0.44 | 108.17 |
| Xiaomi | 45 | 0.26 (−0.08 to 0.54) | 0.15 | −0.08 to 0.39 | 293.29 |
| Fitbit | 298 | 0.56[a] (0.47 to 0.63) | 0.56[a] | 0.48 to 0.64 | 114.30 |

[a]*P*<.001.

Correlations for steps and MVPA are illustrated in Figures 3 and 4. Scatter and deviation of the points around the line that reflects the perfect agreement between the measurements are larger for measuring MVPA than for measuring steps. The largest scatter for measuring steps is found for iWown (Figure 3). On the basis of scatterplots, a careful statement on overestimation or underestimation of the measurement of the trackers can be made. This is based on the location of the data points relative to the line that represents the perfect agreement between the measurements. For Xiaomi, Nokia, and VeryFit, the majority of the data points are located below that line, meaning an underestimation of the number of steps. For iWown, MyKronoz, and Fitbit, the majority of the data points are located above the line, meaning an overestimation of the number of steps. For Geonaut, no clear underestimation or overestimation is visualized. A large scatter for all 3 trackers that measure MVPA was observed, with no obvious relation between the MVPA measurements of the trackers and the MVPA measurements of the ActiGraph. For Nokia, an underestimation is visualized, and for Xiaomi, however, an overestimation is visualized. For Fitbit, no clear underestimation or overestimation is visualized.

These findings are also visualized by using Bland-Altman plots. Bland-Altman plots were used to visualize the differences between the steps and MVPA measurements of the ActiGraph accelerometer and each tracker (y-axis) against the average number of steps or number of minutes of MVPA of the measurements of these 2 devices (x-axis). Mean differences with the ActiGraph accelerometer and the limits of agreement are presented in Table 3 (illustrated in Figures 5 and 6 for steps and MVPA, respectively). A positive value of the mean difference indicates an underestimation of the measurements of the tracker compared with the ActiGraph measurements, whereas a negative value indicates an overestimation. The systematic overestimation or underestimation (mean differences) and the range between the upper and lower limits of the agreement reflect the accuracy of the measurements of the tracker compared with the measurements of the ActiGraph accelerometer. The broader the range between the lower and the upper limit, the less accurate the measurements are.

For measuring steps and MVPA, the table and the plots (Figures 5 and 6) all showed large limits. The Xiaomi tracker showed the narrowest limits (7450 steps) for measuring steps, whereas iWown showed the broadest limits (19,263 steps). These results are in line with the interpretations of validity findings based on the Spearman *r*, the ICC, and the MAPE score.

For MVPA, the ranges between the lower and upper limit of agreement are very large, indicating a low accurate measurement by all 3 trackers measuring MVPA. The Bland-Altman plots showed the broadest limits for Xiaomi (207.64 min) and the narrowest limits for Nokia (101.80 min).

Thus, several but not all low-cost trackers showed high accuracy to measure steps. Xiaomi trackers even outperformed the Fitbit tracker in measuring steps. However, none of the trackers showed good accuracy to measure MVPA, including Fitbit, which did nevertheless reach a slightly higher validity than the low-cost trackers in measuring MVPA.

**Figure 3.** Correlations between steps estimates per day from the trackers and the ActiGraph.

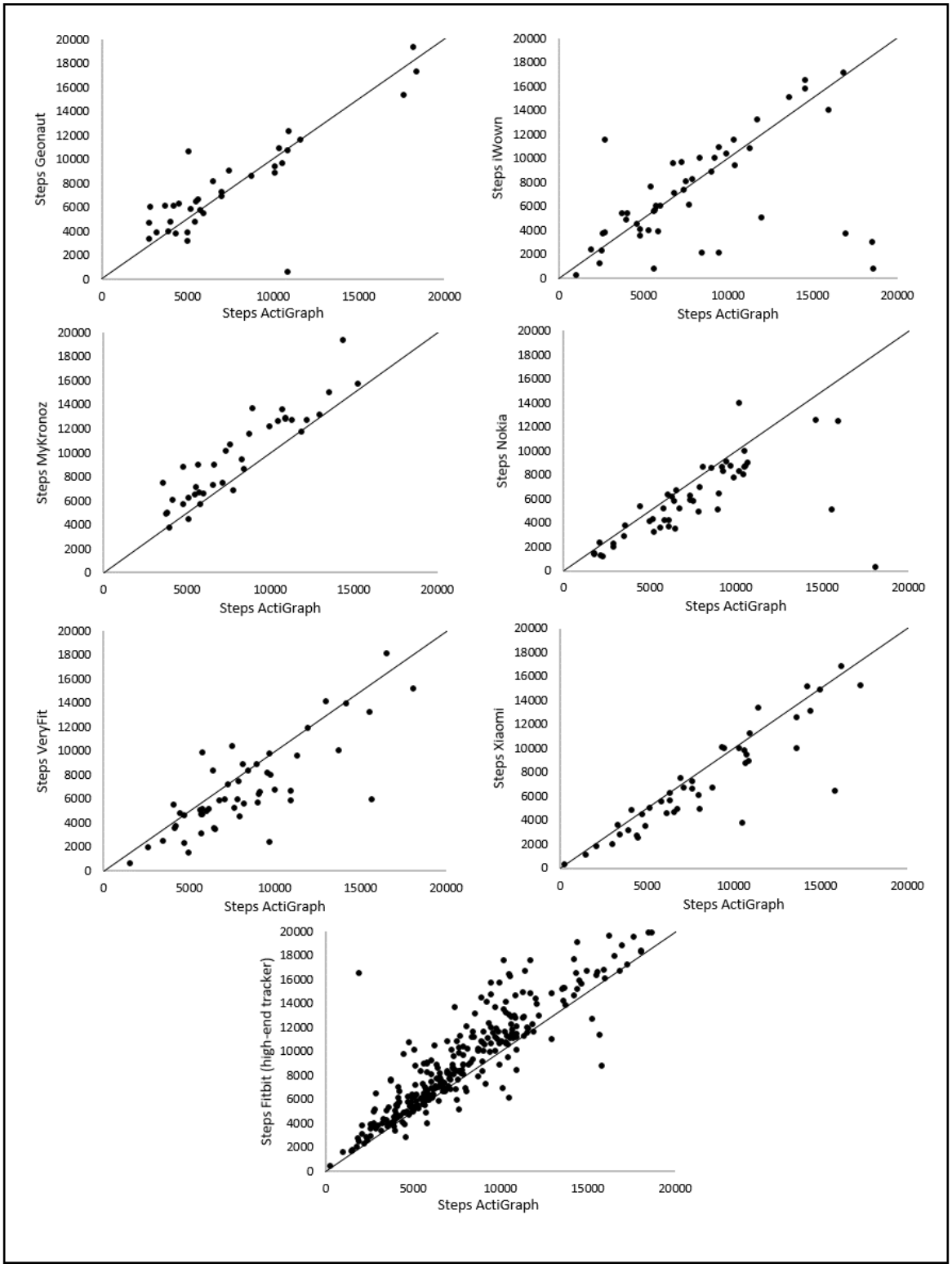


**Figure 4.** Correlations between moderate-to-vigorous physical activity estimates per day from the trackers and the ActiGraph.

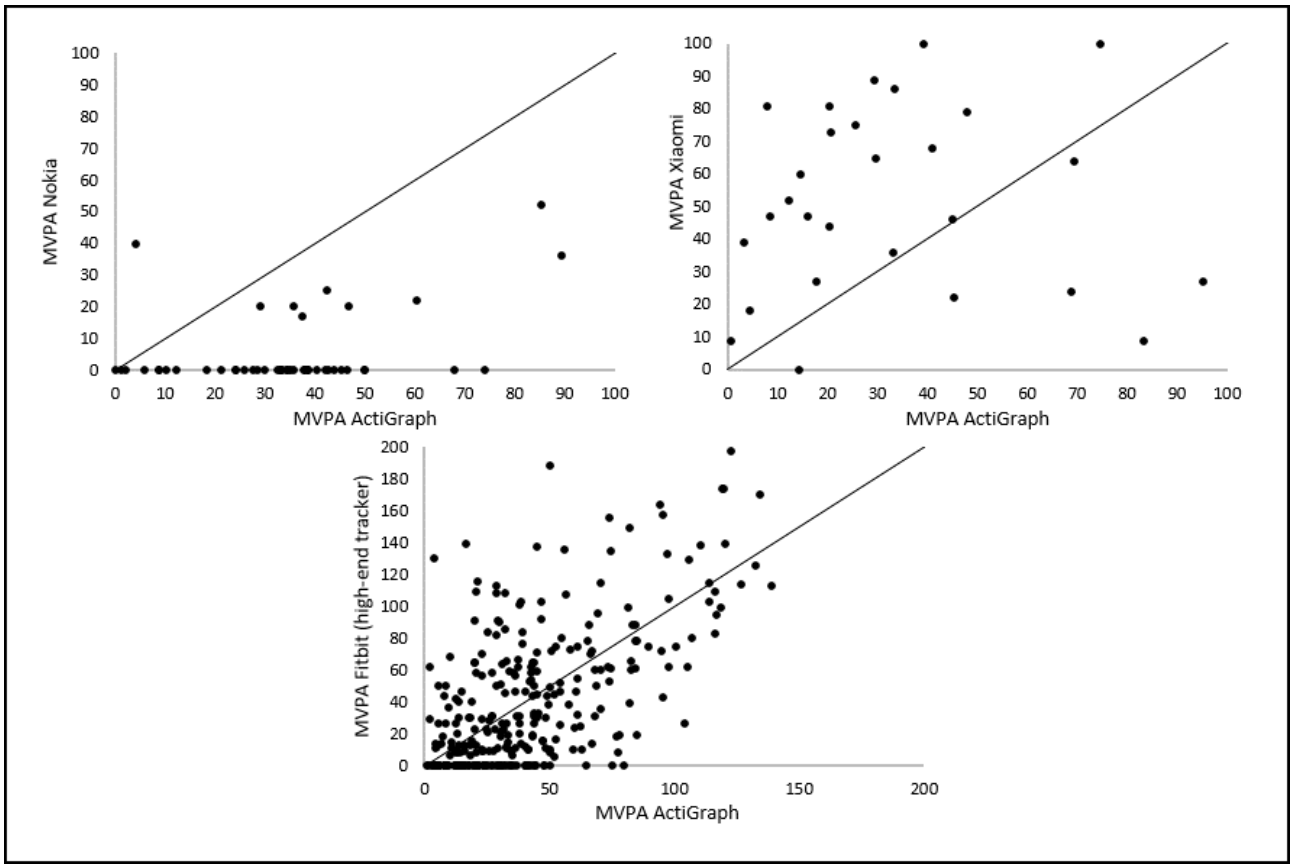



**Table 3.** Mean differences of activity measures with the ActiGraph accelerometer and limits of agreement of the activity trackers.

| Tracker | n | Mean difference of steps (ActiGraph−Tracker) | Limits of agreement, range | Width of the limits of agreement |
|---|---|---|---|---|
| **Steps** | | | | |
| Geonaut | 36 | −146 | −4802 to 4509 | 9311 |
| iWown | 50 | 638 | −8993 to 10,270 | 19,263 |
| MyKronoz | 38 | −1798 | −5563 to 1967 | 7530 |
| Nokia | 50 | 1609 | −4229 to 7447 | 11,676 |
| VeryFit | 54 | 1356 | −3276 to 5989 | 9265 |
| Xiaomi | 45 | 1011 | −2713 to 4737 | 7450 |
| Fitbit | 300 | −1369 | −5238 to 2499 | 7737 |
| **Moderate-to-vigorous physical activity** | | | | |
| Nokia | 16 | 32.55 | −18.35 to 83.45 | 101.80 |
| Xiaomi | 45 | −35.14 | −138.96 to 68.68 | 207.64 |
| Fitbit | 298 | −1.27 | −77.07 to 74.52 | 151.59 |

**Figure 5.** Bland-Altman plots of the trackers for measuring steps. The middle line shows the mean difference (positive values indicate an underestimation of the wearable and negative values indicate an overestimation) between the measurements of steps of the wearables and the ActiGraph and the dashed lines indicate the limits of agreement (1.96×SD of the difference scores).

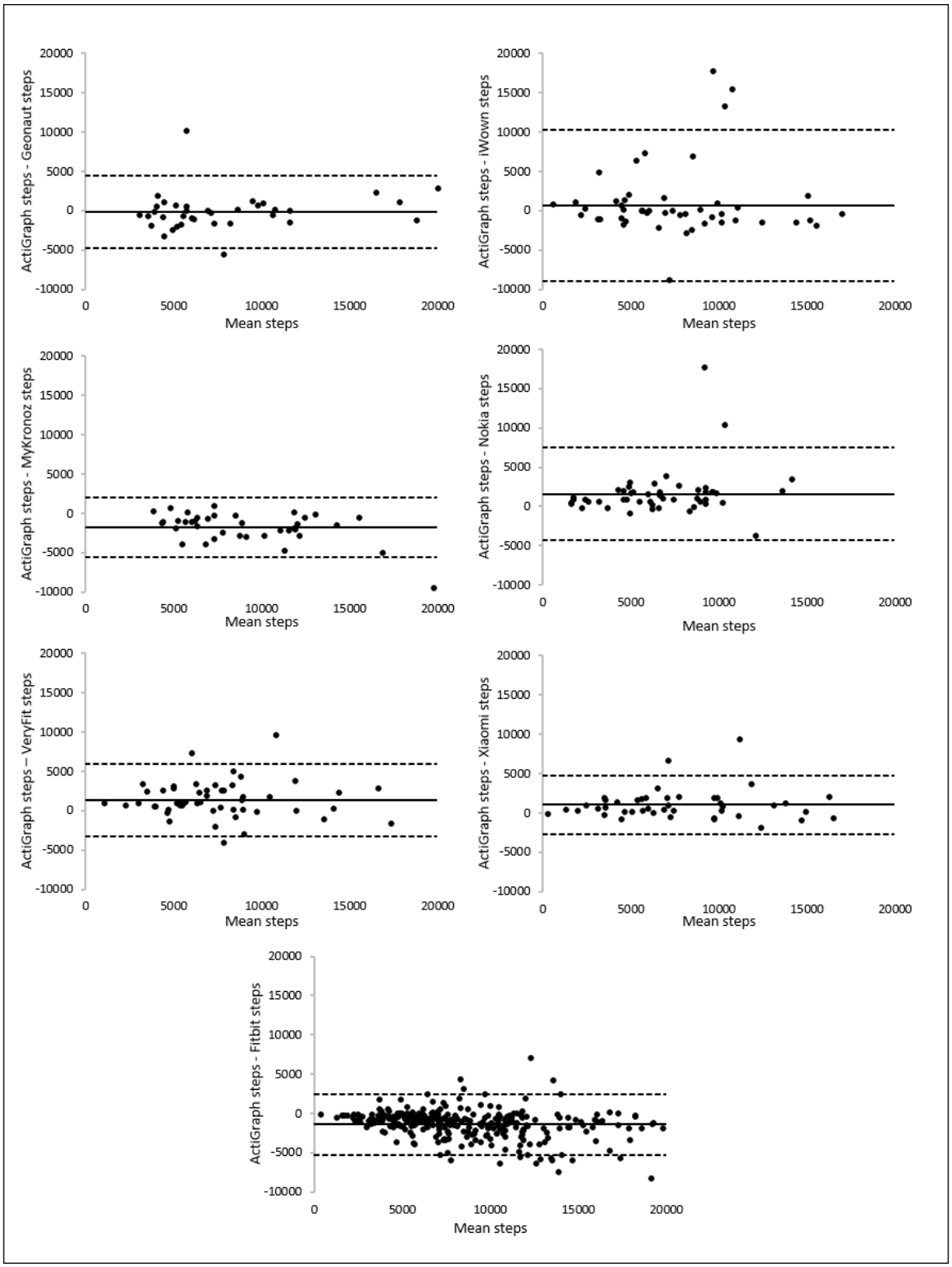
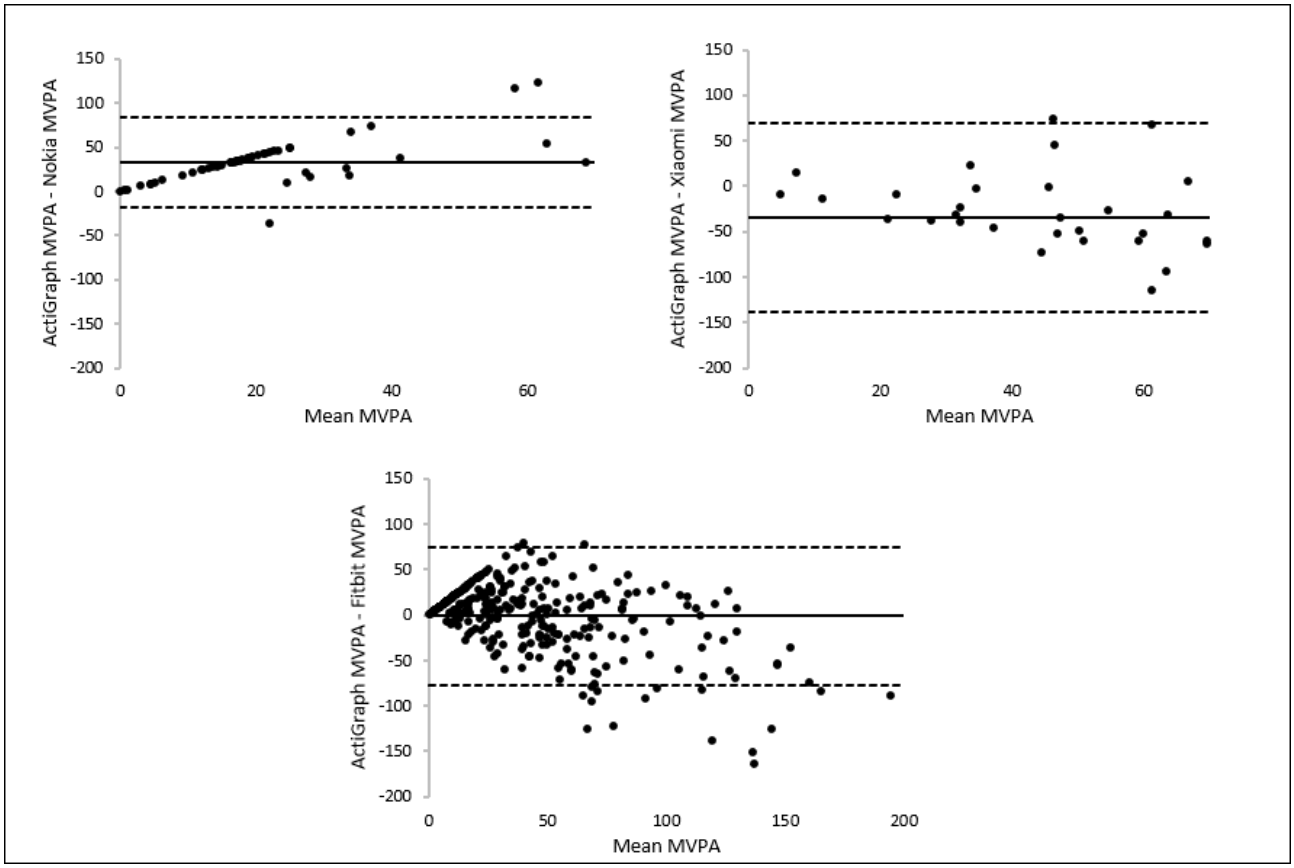


**Figure 6.** Bland-Altman plots of the trackers for measuring moderate-to-vigorous physical activity. The middle line shows the mean difference (positive values indicate an underestimation of the wearable, and negative values indicate an overestimation) between the measurements of moderate-to-vigorous physical activity of the wearables and the ActiGraph, and the dashed lines indicate the limits of agreement (1.96×SD of the difference scores).




### *Total Sleep Time*

Table 4 reports the mean minutes of TST and corresponding standard deviations for all trackers.

Spearman correlations between the TST measurements of the trackers and the TST measurements of the SenseWear armband show large diversity between trackers, ranging from very weak (Geonaut) to strong (VeryFit). The ICCs, however, indicate low agreement (ICC<0.60) between the measurements of all trackers and the measurements of the SenseWear. This could reflect a systematic underestimation or overestimation of TST by the trackers, which is not evident from the Spearman *r* coefficient. The MAPE scores of all trackers also indicate a large mean deviation from the SenseWear measurements for TST, ranging from 20.57% for Fitbit to 39.08% for Xiaomi. The correlation coefficients, ICC values, associated 95% CI, and MAPE scores for measuring TST are shown in Table 5.

**Table 4.** Mean total sleep time per day measured by the low-cost trackers, Fitbit and SenseWear.

| Tracker | Number of measured days | Total sleep time (minutes), mean (SD) | Range (minutes) |
|---|---|---|---|
| **Total sleep time** | | | |
| Geonaut | 30 | 341 (123) | 110-589 |
| iWown | 52 | 421 (108) | 91-624 |
| MyKronoz | 24 | 457 (143) | 70-746 |
| Nokia | 46 | 464 (108) | 247-743 |
| VeryFit | 51 | 472 (59) | 193-614 |
| Xiaomi | 44 | 495 (87) | 285-695 |
| Fitbit | 287 | 414 (91) | 68-733 |
| SenseWear | 147 | 373 (83) | 112-653 |

**Table 5.** Correlation coefficients, intraclass correlation coefficients, associated 95% CI of the measurements, and mean absolute percentage error scores for measuring total sleep time.

| Tracker | n | Spearman *r* (95% CI) | Intraclass correlation coefficient | 95% CI | Mean absolute percentage error (%) |
|---|---|---|---|---|---|
| **Total sleep time** | | | | | |
| Geonaut | 15 | 0.04 (−0.45 to 0.60) | 0.05 | −0.44 to 0.52 | 26.59 |
| iWown | 24 | 0.57[a] (0.19 to 0.84) | 0.52[a] | 0.18 to 0.76 | 21.33 |
| MyKronoz | 14 | 0.45 (−0.22 to 0.86) | 0.40[a] | −0.07 to 0.74 | 38.15 |
| Nokia | 14 | 0.66[b] (0.30 to 0.88) | 0.30[a] | −0.10 to 0.63 | 38.63 |
| VeryFit | 24 | 0.73[b] (0.48 to 0.83) | 0.26 | −0.11 to 0.61 | 30.73 |
| Xiaomi | 21 | 0.21 (−0.34 to 0.68) | 0.13 | −0.13 to 0.45 | 39.08 |
| Fitbit | 134 | 0.57[b] (0.40 to 0.69) | 0.46[b] | 0.28 to 0.60 | 20.57 |

[a]*P*<.05.

[b]*P*<.001.

The correlations for TST are also illustrated in Figure 7. This figure visualizes the large discrepancy between the Spearman correlation coefficient and the ICC, specifically evident for Nokia and VeryFit. Although a clear relation is visible between the measurements (Spearman *r*), almost all data points are above the line that represents the perfect agreement between the measurements. This indicates a systematic overestimation of the TST measurements of Nokia and VeryFit compared with the convergent measure. Figure 7 also shows the largest scatter for MyKronoz.

Bland-Altman plots for TST revealed the smallest limits for VeryFit (263.39 min) and the broadest limits for Geonaut (558.25 min). These results are in line with the findings based on the Spearman *r* coefficient and the scatter of the data points. The mean differences with the SenseWear armband measurements and the limits of agreement are presented in Table 6 and illustrated in Figure 8.

**Figure 7.** Correlations between total sleep time estimates from the trackers and the SenseWear.

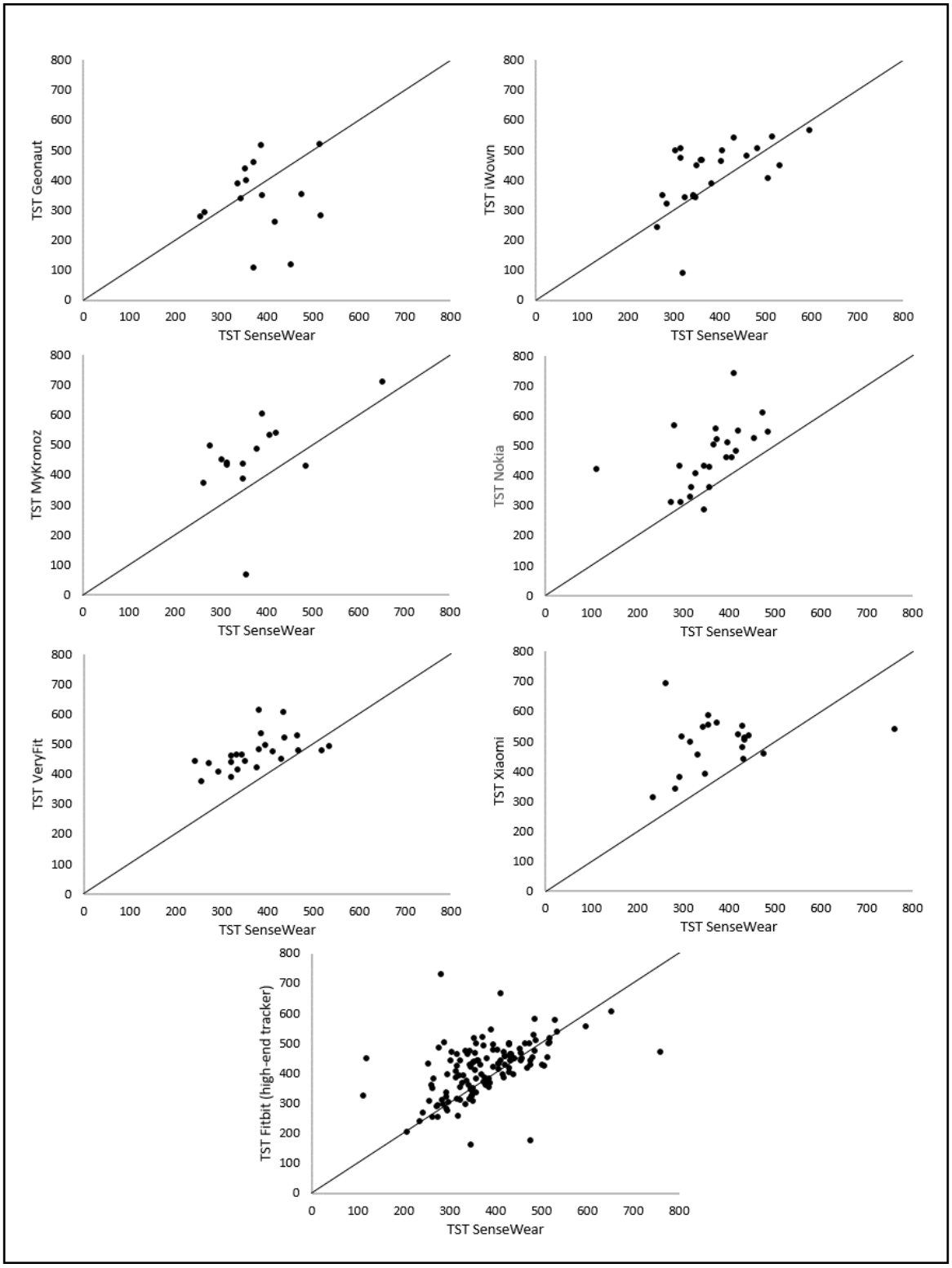
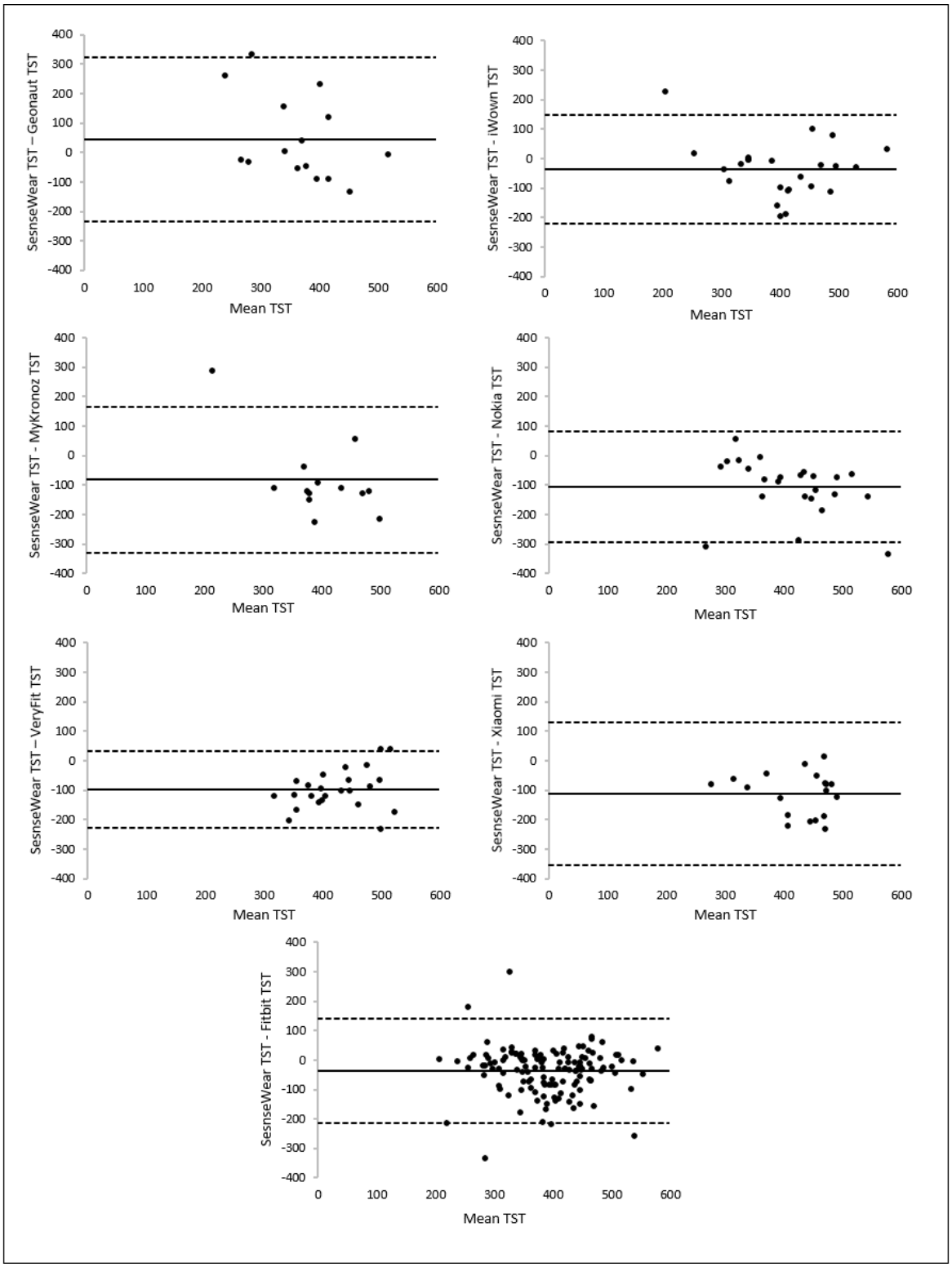


**Table 6.** Mean differences of total sleep time measures with the SenseWear and limits of agreement of the activity trackers.

| Tracker | n | Mean difference of total sleep time (SenseWear−smartwatch) | Limits of agreement, range | Width of the limits of agreement (minutes) |
|---|---|---|---|---|
| Geonaut | 15 | 44.93 | −234.19 to 324.06 | 558.25 |
| iWown | 24 | −36.79 | −221.22 to 147.63 | 368.85 |
| MyKronoz | 14 | −82.29 | −330.55 to 165.98 | 496.53 |
| Nokia | 24 | −106.46 | −293.72 to 80.80 | 374.52 |
| VeryFit | 24 | −97.63 | −229.32 to 34.07 | 263.39 |
| Xiaomi | 21 | −112.14 | −355.40 to 131.12 | 486.52 |
| Fitbit | 134 | −36.91 | −213.9 to 140.16 | 354.14 |

**Figure 8.** Bland-Altman plots of the trackers for measuring total sleep time. The middle line shows the mean difference (positive values indicate an underestimation of the wearable and negative values indicate an overestimation) between the measurements of total sleep time of the wearables and the SenseWear, and the dashed lines indicate the limits of agreement (1.96×SD of the difference scores).




Thus, low-cost trackers showed low (eg, Geonaut and Xiaomi) to strong (eg, VeryFit) correlations to measure TST, with some trackers such as VeryFit and Nokia systematically overestimating TST. Fitbit showed low (based on ICC) to moderate (based on the Spearman *r* coefficient) validity to measure TST and was outperformed by VeryFit to measure TST on all indicators of accuracy.

## Discussion

### Principal Findings

This study examined the validity of low-cost trackers (≤€50 [US $56]) for measuring adults' steps, moderate-to-vigorous PA, and TST in free-living conditions. In general, the low-cost trackers were most accurate in the measurement of steps,

somewhat accurate for the measurement of sleep, and lacked validity for the measurement of MVPA time. Validity ranged widely between the various low-cost trackers tested. The performance of the best of the low-cost trackers approached or even exceeded that of the Fitbit Charge 2 (the high-cost comparison tracker), whereas the worst of the low-cost trackers had weak validity. Notably, VeryFit 2.0 performed relatively strongly across both sleep and steps domains, whereas the Xiaomi Mi Band 2 appeared to have the highest validity for the measurement of steps.

The finding that many of the low-cost trackers are accurate for measuring steps is promising, given that steps is the metric reported by users of trackers as being of most interest [63]. We found that the low-cost trackers were most accurate for measuring steps in comparison with sleep and minutes of MVPA. This order for validity (ie, measuring steps more accurately than measuring sleep and in turn more accurately than measuring MVPA) is consistent with findings for these metrics in high-cost trackers [39], although in our study, the low-cost trackers demonstrated weak-to-moderate validity for MVPA minutes (Spearman *r* ranged from 0.24 to 0.56), whereas previous research in high-cost trackers has suggested moderate-to-strong validity (eg, Ferguson et al's [39] study of high-cost trackers reported Pearson *r* ranging from 0.52 to 0.91). It is possible that some of the differences between the reference values for MVPA derived from the ActiGraph accelerometers and the values recorded by the low-cost trackers may have originated from a measurement error associated with the reference device. Furthermore, a possible explanation for the weak-to-nil validity found in our study could be that the PA variables measured by the low-cost trackers were not explicitly identified as MVPA. However, because all devices had set a goal of 30-min PA per day (similar to the MVPA recommendations for adults), we assumed that the measured variable corresponded to MVPA as measured by the ActiGraph accelerometer. Nevertheless, specific information regarding algorithm intensity cut-points was not provided and publicly available from these low-cost trackers. Therefore, the discrepancies in this study may be a result of both definitional and measurement problems (eg, sensitivity algorithm). In this regard, it may be very useful in the future, when manufacturers provide more insight into the cut-points and algorithms that were used to translate the raw data into useful information (such as steps and minutes of MVPA).

Although research-grade accelerometers are the closest we have to a *gold standard* for the measurement of MVPA in free-living conditions, the MVPA values derived from them can vary by the order of magnitude depending on parameters such as epoch length and cut-points [64]. Furthermore, wear position has an impact on the validity of MVPA. Studies comparing the validity of research-grade accelerometers at different body locations consistently show that the hip position is more accurate than the wrist [65]. Despite the recognized superior validity of hip-worn accelerometers and trackers, over the past 5 years or so, there has been a shift for both consumer trackers and research-grade accelerometers to increasingly be designed for wrist wear, presumably because of improved logistics, such as comfort and convenience. This clear shift in the market highlights that validity should not be considered the be-all and end-all. Issues such as usability, compliance, and adherence are also important, although they tend to receive less attention in the scientific literature.

Evidence for the validity of the low-cost trackers for the measurement of sleep duration was mixed. Some trackers performed quite strongly. For example, the top performing tracker, VeryFit 2.0, demonstrated a Spearman *r* of 0.73 for TST compared with the reference device (SenseWear), which was actually superior to the Fitbit Charge HR ($r$=0.57). However, the Bland-Altman analyses revealed that VeryFit 2.0 tended to overestimate sleep by around 1.5 hours per night compared with the reference device. If this overestimation was consistent, it could be argued that the data might still be useful for self-monitoring changes in sleep over time. However, the Bland-Altman 95% limits of agreement spanned a range of 263 min, suggesting that the extent of overestimation varied considerably on different administrations. It, therefore, seems questionable whether the sleep estimates derived from VeryFit 2.0 are accurate enough to help a user meaningfully monitor/change their sleeping patterns.

The finding that low-cost trackers have strong validity for measuring steps and some validity for measuring sleep is likely to be of interest to public health researchers and clinicians alike. There is considerable interest in using activity trackers to intervene on lifestyle activities, with a recent meta-analysis finding positive evidence for short-term effectiveness but less evidence for sustained effects [66]. There is well-recognized usage attrition associated with activity trackers over time. For example, a 2017 study gave entry-level Fitbit trackers to 711 users and found that approximately 50% of participants had stopped using them at 6 months and 80% had stopped by 10 months [67]. The most common reasons for not using Fitbit was technical failure or difficulty (57%), losing the device (13%), or forgetting to wear it (13%). Nonetheless, low-cost devices fill an important gap in the consumer market between the high-cost activity trackers that are prohibitively expensive to provide to clinical or research cohorts at scale (unless sizable funding is available) but likely to be more aesthetically pleasing and acceptable to wearers than traditional pedometers [19,63]. The findings of this study, which highlight the Xiaomi Mi Band 2 and VeryFit 2.0 devices as having acceptable validity, are therefore helpful. We bought the trackers as individual buyers on the consumer market. Researchers intending to use these in large-scale research cohorts may purchase these at an even lower cost in bulk. Another promising feature of VeryFit 2.0 is that it has an application programming interface (API) that allows software developers to create custom software that can be integrated directly with the tracker (ie, data from the tracker can be sent automatically to the custom software). There is a growing trend for eHealth and mobile health research to use Fitbit and Garmin API [68-70]. Therefore, validated low-cost trackers with APIs offer new data collection and intervention possibilities.

Our study included trackers with and without heart rate measurements. All trackers with the highest validity included heart rate measures, whereas those without showed lower validity. However, we cannot conclude from this study that the

heart rate function increased validity. Studies testing the same model with and without the heart rate function and assessing the validity of the heart rate measurement in itself would be needed to make this claim. The price of included trackers ranged from around €25 (US $28) to approximately €50 (US $55). The prices of the most accurate types, VeryFit 2.0 and Xiaomi Mi Band 2, are situated in the middle of this range (€30-€40 [US $33-US $44]). This yields two models that are very attractive and accessible to the general public. Thus, price may not be the determining factor in the validity of the trackers: more expensive within this range is not necessarily better. On the contrary, we cannot conclude that price plays no role and that trackers even less expensive than those included here (<€25 [US $28]) may also be valid. Indeed, a study on pedometers provided for free as gadgets with cereal boxes found that those were not valid [71].

Although validity evidence from this study for low-cost devices measuring steps, MVPA, and TST is not unequivocally good across the devices, user experience is also extremely important. A device that has high validity may not necessarily have a positive user experience. Future research examining the user experience of low-cost trackers (eg, focusing on issues such as functionality, reliability, and ease of use, both of the device itself and its accompanying app) will be valuable. Our preliminary experiences suggest that the user experience of the low-cost trackers may be less positive than that of the high-cost trackers (eg, we tended to experience fewer technical issues with Fitbit trackers than with the other devices in this study). It can be assumed that the higher price of the high-cost trackers is partly determined by the investments made by the manufacturer to improve the user experience and to better develop the app supporting the tracker. Moreover, the low-cost activity trackers appear most valid for measuring steps. Pedometers that count steps are available at an even lower cost, but unlike activity trackers, they offer little additional functionality (eg, feedback, information, and social support) in an accompanying app and are considered less usable by people than activity trackers [72-74]. Further work to explore these issues more rigorously and in greater depth is warranted.

### Strengths and Limitations

A strength of our study is that it is the first to scrutinize the validity of low-cost trackers addressing an important gap in the scientific literature to date. Methodological strengths of the study are the relatively large number of devices that were tested using the same methodology (allowing a direct comparison of the devices' performance), the devices that tested multiple metrics (steps, MVPA, and sleep), and that efforts were made to minimize bias, for example, by randomizing the order in which participants wore the devices. Limitations included that our sample was relatively young and healthy. On the basis of previous literature, it seems likely that validity for measuring steps is likely to be somewhat lower in older and clinical populations (eg, obese) [75]. As already noted, our reference devices were research-grade accelerometers with known validity limitations of their own. Therefore, they represent convergent validity rather than criterion validity, and there is a risk that we may be underestimating the low-cost trackers' true validity. A further limitation is that this is a fast-moving field with new devices continually entering and exiting the market. In particular, since our study started, the Xiaomi MiBand 2 is replaced by its successor, the MiBand 3. Therefore, it would be beneficial that future research continuously investigates the validity of new low-cost trackers and other emerging devices. Furthermore, having an insight into the used algorithms and used cutoffs would be beneficial.

### Conclusions

This study was the first to examine the validity of low-cost trackers. It found that validity was strongest for the measurement of steps; there was some evidence of validity for the measurement of sleep, whereas validity for the measurement of MVPA time was weak. Validity ranged between devices, with Xiaomi having the highest validity for the measurement of steps and VeryFit performing relatively strongly across both sleep and steps domains. The tested low-cost trackers hold promise for the cost-efficient measurement of movement behaviors. Further research investigating the user experience of low-cost devices and their accompanying apps is needed before these devices can be confidently recommended.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Parameter estimates from linear mixed effects models examining the association between commercial trackers and ActiGraph (steps, moderate-to-vigorous physical activity) and SenseWear (total sleep time).
[DOCX File , 15 KB - mhealth_v8i5e16674_app1.docx ]

## References

1. Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. J Appl Physiol (1985) 2005 Sep;99(3):1193-1204 [FREE Full text] [doi: 10.1152/japplphysiol.00160.2005] [Medline: 16103522]
2. Cappuccio FP, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J 2011 Jun;32(12):1484-1492. [doi: 10.1093/eurheartj/ehr007] [Medline: 21300732]

3. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010 Feb;33(2):414-420 [FREE Full text] [doi: 10.2337/dc09-1124] [Medline: 19910503]
4. Guallar-Castillón P, Bayán-Bravo A, León-Muñoz LM, Balboa-Castillo T, López-García E, Gutierrez-Fisac JL, et al. The association of major patterns of physical activity, sedentary behavior and sleep with health-related quality of life: a cohort study. Prev Med 2014 Oct;67:248-254. [doi: 10.1016/j.ypmed.2014.08.015] [Medline: 25138382]
5. Magee L, Hale L. Longitudinal associations between sleep duration and subsequent weight gain: a systematic review. Sleep Med Rev 2012 Jun;16(3):231-241 [FREE Full text] [doi: 10.1016/j.smrv.2011.05.005] [Medline: 21784678]
6. Maher CA, Mire E, Harrington DM, Staiano AE, Katzmarzyk PT. The independent and combined associations of physical activity and sedentary behavior with obesity in adults: NHANES 2003-06. Obesity (Silver Spring) 2013 Dec;21(12):E730-E737 [FREE Full text] [doi: 10.1002/oby.20430] [Medline: 23512825]
7. Cassidy S, Chau JY, Catt M, Bauman A, Trenell MI. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233,110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. BMJ Open 2016 Mar 15;6(3):e010038 [FREE Full text] [doi: 10.1136/bmjopen-2015-010038] [Medline: 27008686]
8. Chastin SF, Palarea-Albaladejo J, Dontje ML, Skelton DA. Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: a novel compositional data analysis approach. PLoS One 2015;10(10):e0139984 [FREE Full text] [doi: 10.1371/journal.pone.0139984] [Medline: 26461112]
9. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]
10. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
11. Middelweerd A, van der Laan DM, van Stralen MM, Mollee JS, Stuij M, te Velde SJ, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 1;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
12. DeSmet A, de Bourdeaudhuij I, Chastin S, Crombez G, Maddison R, Cardon G. Adults' preferences for behavior change techniques and engagement features in a mobile app to promote 24-hour movement behaviors: cross-sectional survey study. JMIR Mhealth Uhealth 2019 Dec 20;7(12):e15707 [FREE Full text] [doi: 10.2196/15707] [Medline: 31859680]
13. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
14. Hickey AM, Freedson PS. Utility of consumer physical activity trackers as an intervention tool in cardiovascular disease prevention and treatment. Prog Cardiovasc Dis 2016;58(6):613-619. [doi: 10.1016/j.pcad.2016.02.006] [Medline: 26943981]
15. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008 Nov;19(6):838-845 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7b0] [Medline: 18854708]
16. Gomersall SR, Ng N, Burton NW, Pavey TG, Gilson ND, Brown WJ. Estimating physical activity and sedentary behavior in a free-living context: a pragmatic comparison of consumer-based activity trackers and ActiGraph accelerometry. J Med Internet Res 2016 Sep 7;18(9):e239 [FREE Full text] [doi: 10.2196/jmir.5531] [Medline: 27604226]
17. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
18. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
19. Alley S, Schoeppe S, Guertler D, Jennings C, Duncan MJ, Vandelanotte C. Interest and preferences for using advanced physical activity tracking devices: results of a national cross-sectional survey. BMJ Open 2016 Jul 7;6(7):e011243 [FREE Full text] [doi: 10.1136/bmjopen-2016-011243] [Medline: 27388359]
20. Vandelanotte C, Duncan MJ, Maher CA, Schoeppe S, Rebar AL, Power DA, et al. The effectiveness of a web-based computer-tailored physical activity intervention using Fitbit activity trackers: randomized trial. J Med Internet Res 2018 Dec 18;20(12):e11321 [FREE Full text] [doi: 10.2196/11321] [Medline: 30563808]
21. Olsen HM, Brown WJ, Kolbe-Alexander T, Burton NW. A brief self-directed intervention to reduce office employees' sedentary behavior in a flexible workplace. J Occup Environ Med 2018 Oct;60(10):954-959. [doi: 10.1097/JOM.0000000000001389] [Medline: 30001255]
22. Vandendriessche K, de Marez L. Imec R&D. Leuven: imec; 2019. [Digital Media Trends in Flanders] URL: https://www.imec.be/static/476531-IMEC-Digimeter-Rapport%202020-WEB-f6fe6fa9efafc16bd174620ca1f90376.PDF?utm_source=website&utm_medium=button&utm_campaign=digimeter_rapport_2019 [accessed 2019-08-17]
23. Jia Y, Wang W, Wen D, Liang L, Gao L, Lei J. Perceived user preferences and usability evaluation of mainstream wearable devices for health monitoring. PeerJ 2018;6:e5350 [FREE Full text] [doi: 10.7717/peerj.5350] [Medline: 30065893]

24. Beenackers MA, Kamphuis CB, Giskes K, Brug J, Kunst AE, Burdorf A, et al. Socioeconomic inequalities in occupational, leisure-time, and transport related physical activity among European adults: a systematic review. Int J Behav Nutr Phys Act 2012 Sep 19;9:116 [FREE Full text] [doi: 10.1186/1479-5868-9-116] [Medline: 22992350]
25. Lallukka T, Sares-Jäske L, Kronholm E, Sääksjärvi K, Lundqvist A, Partonen T, et al. Sociodemographic and socioeconomic differences in sleep duration and insomnia-related symptoms in Finnish adults. BMC Public Health 2012 Jul 28;12:565 [FREE Full text] [doi: 10.1186/1471-2458-12-565] [Medline: 22839359]
26. Murray E, Hekler E, Andersson G, Collins L, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
27. Sørensen K, van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
28. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of Health 1.0 and Health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
29. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
30. Porter CE, Donthu N. Using the technology acceptance model to explain how attitudes determine internet usage: The role of perceived access barriers and demographics. J Bus Res 2006 Sep;59(9):999-1007. [doi: 10.1016/j.jbusres.2006.06.003]
31. 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]
32. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 5;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
33. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
34. Tully MA, McBride C, Heron L, Hunter RF. The validation of Fibit Zip physical activity monitor as a measure of free-living physical activity. BMC Res Notes 2014 Dec 23;7:952 [FREE Full text] [doi: 10.1186/1756-0500-7-952] [Medline: 25539733]
35. Sasaki JE, Hickey A, Mavilia M, Tedesco J, John D, Keadle SK, et al. Validation of the Fitbit wireless activity tracker for prediction of energy expenditure. J Phys Act Health 2015 Feb;12(2):149-154. [doi: 10.1123/jpah.2012-0495] [Medline: 24770438]
36. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
37. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
38. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
39. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015 Mar 27;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
40. Schneider M, Chau L. Validation of the Fitbit Zip for monitoring physical activity among free-living adolescents. BMC Res Notes 2016 Sep 21;9(1):448 [FREE Full text] [doi: 10.1186/s13104-016-2253-6] [Medline: 27655477]
41. Mantua J, Gravel N, Spencer R. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 5;16(5):pii: E646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
42. de Zambotti M, Baker FC, Willoughby AR, Godino JG, Wing D, Patrick K, et al. Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. Physiol Behav 2016 May 1;158:143-149 [FREE Full text] [doi: 10.1016/j.physbeh.2016.03.006] [Medline: 26969518]
43. Dickinson DL, Cazier J, Cech T. A practical validation study of a commercial accelerometer using good and poor sleepers. Health Psychol Open 2016 Jul;3(2):2055102916679012 [FREE Full text] [doi: 10.1177/2055102916679012] [Medline: 28815052]
44. Brooke SM, An H, Kang S, Noble JM, Berg KE, Lee J. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]

45. 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 (1985) 2004 Jan;96(1):343-351 [FREE Full text] [doi: 10.1152/japplphysiol.00703.2003] [Medline: 12972441]
46. van de Water AT, Holmes A, Hurley D. Objective measurements of sleep for non-laboratory settings as alternatives to polysomnography--a systematic review. J Sleep Res 2011 Mar;20(1 Pt 2):183-200 [FREE Full text] [doi: 10.1111/j.1365-2869.2009.00814.x] [Medline: 20374444]
47. Colon A, Duffy J. PCMag. 2020 Feb 13. The Best Fitness Trackers for 2020 URL: https://www.pcmag.com/picks/the-best-fitness-trackers [accessed 2019-05-17]
48. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 2017;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
49. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
50. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 4;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
51. Rowlands AV, Stone MR, Eston RG. Influence of speed and step frequency during walking and running on motion sensor output. Med Sci Sports Exerc 2007 Apr;39(4):716-727. [doi: 10.1249/mss.0b013e318031126c] [Medline: 17414811]
52. Le Masurier GC, Lee SM, Tudor-Locke C. Motion sensor accuracy under controlled and free-living conditions. Med Sci Sports Exerc 2004 May;36(5):905-910. [doi: 10.1249/01.mss.0000126777.50188.73] [Medline: 15126728]
53. Le Masurier GC, Tudor-Locke C. Comparison of pedometer and accelerometer accuracy under controlled conditions. Med Sci Sports Exerc 2003 May;35(5):867-871. [doi: 10.1249/01.MSS.0000064996.63632.10] [Medline: 12750599]
54. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc accelerometer. Med Sci Sports Exerc 1998 May;30(5):777-781. [doi: 10.1097/00005768-199805000-00021] [Medline: 9588623]
55. Santos-Lozano A, Santín-Medeiros F, Cardon G, Torres-Luque G, Bailón R, Bergmeir C, et al. Actigraph GT3X: validation and determination of physical activity intensity cut points. Int J Sports Med 2013 Nov;34(11):975-982. [doi: 10.1055/s-0033-1337945] [Medline: 23700330]
56. Rich C, Geraci M, Griffiths L, Sera F, Dezateux C, Cortina-Borja M. Quality control methods in accelerometer data processing: defining minimum wear time. PLoS One 2013;8(6):e67206 [FREE Full text] [doi: 10.1371/journal.pone.0067206] [Medline: 23826236]
57. Troiano RP, Berrigan D, Dodd KW, Mâsse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc 2008 Jan;40(1):181-188. [doi: 10.1249/mss.0b013e31815a51b3] [Medline: 18091006]
58. Shin M, Swan P, Chow CM. The validity of Actiwatch2 and SenseWear armband compared against polysomnography at different ambient temperature conditions. Sleep Sci 2015;8(1):9-15 [FREE Full text] [doi: 10.1016/j.slsci.2015.02.003] [Medline: 26483937]
59. Malavolti M, Pietrobelli A, Dugoni M, Poli M, Romagnoli E, de Cristofaro P, et al. A new device for measuring resting energy expenditure (REE) in healthy subjects. Nutr Metab Cardiovasc Dis 2007 Jun;17(5):338-343. [doi: 10.1016/j.numecd.2005.12.009] [Medline: 17562571]
60. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
61. Kim Y, Park I, Kang M. Convergent validity of the international physical activity questionnaire (IPAQ): meta-analysis. Public Health Nutr 2013 Mar;16(3):440-452. [doi: 10.1017/S1368980012002996] [Medline: 22874087]
62. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
63. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017 Nov 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]
64. Banda JA, Haydel KF, Davila T, Desai M, Bryson S, Haskell WL, et al. Effects of varying epoch lengths, wear time algorithms, and activity cut-points on estimates of child sedentary behavior and physical activity from accelerometer data. PLoS One 2016;11(3):e0150534 [FREE Full text] [doi: 10.1371/journal.pone.0150534] [Medline: 26938240]
65. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. Med Sci Sports Exerc 2013 May;45(5):964-975 [FREE Full text] [doi: 10.1249/MSS.0b013e31827f0d9c] [Medline: 23247702]
66. Brickwood KJ, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
67. Hermsen S, Moons J, Kerkhof P, Wiekens C, de Groot M. Determinants for sustained use of an activity tracker: observational study. JMIR Mhealth Uhealth 2017 Oct 30;5(10):e164 [FREE Full text] [doi: 10.2196/mhealth.7311] [Medline: 29084709]

68. Shaw R, Levine E, Streicher M, Strawbridge E, Gierisch J, Pendergast J, et al. Log2Lose: development and lessons learned from a mobile technology weight loss intervention. JMIR Mhealth Uhealth 2019 Feb 13;7(2):e11972 [FREE Full text] [doi: 10.2196/11972] [Medline: 30758297]
69. Krans M, van de Wiele L, Bullen N, Diamond M, van Dantzig S, de Ruyter B, et al. A group intervention to improve physical activity at the workplace. In: Oinas-Kukkonen H, Win K, Karapanos E, Karppinen P, Kyza E, editors. Persuasive Technology: Development of Persuasive and Behavior Change Support Systems. Cham: Springer; 2019:322-333.
70. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: The ACTIVATE Trial. Cancer 2019 Aug 15;125(16):2846-2855. [doi: 10.1002/cncr.32143] [Medline: 31012970]
71. de Cocker K, Cardon G, de Bourdeaudhuij I. Validity of the inexpensive Stepping Meter in counting steps in free living conditions: a pilot study. Br J Sports Med 2006 Aug;40(8):714-716 [FREE Full text] [doi: 10.1136/bjsm.2005.025296] [Medline: 16790485]
72. Husted HM, Llewellyn TL. The accuracy of pedometers in measuring walking steps on a treadmill in college students. Int J Exerc Sci 2017;10(1):146-153 [FREE Full text] [Medline: 28479955]
73. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
74. Henriksen A, Mikalsen MH, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, et al. Using fitness trackers and smartwatches to measure physical activity in research: analysis of consumer wrist-worn wearables. J Med Internet Res 2018 Mar 22;20(3):e110 [FREE Full text] [doi: 10.2196/jmir.9157] [Medline: 29567635]
75. Kenyon A, McEvoy M, Sprod J, Maher C. Validity of pedometers in people with physical disabilities: a systematic review. Arch Phys Med Rehabil 2013 Jun;94(6):1161-1170. [doi: 10.1016/j.apmr.2012.11.030] [Medline: 23201318]


## Abbreviations

**API:** application programming interface
**eHealth:** electronic health
**ICC:** intraclass correlation coefficient
**MAPE:** mean absolute percentage error
**MVPA:** moderate-to-vigorous physical activity
**PA:** physical activity
**SB:** sedentary behavior
**SES:** socioeconomic status
**TST:** total sleep time


*Edited by G Eysenbach; submitted 14.10.19; peer-reviewed by J Di, A Kononova; comments to author 23.12.19; revised version received 24.02.20; accepted 26.02.20; published 19.05.20.*

*Please cite as:*
*Degroote L, Hamerlinck G, Poels K, Maher C, Crombez G, De Bourdeaudhuij I, Vandendriessche A, Curtis RG, DeSmet A*
*Low-Cost Consumer-Based Trackers to Measure Physical Activity and Sleep Duration Among Adults in Free-Living Conditions: Validation Study*
*JMIR Mhealth Uhealth 2020;8(5):e16674*
*URL: http://mhealth.jmir.org/2020/5/e16674/*
*doi:10.2196/16674*
*PMID:32282332*

Original Paper

# Augmented Reality–Based Rehabilitation of Gait Impairments: Case Report

Jeremia Philipp Oskar Held[1*], MSc, PhD; Kevin Yu[2*], MSc; Connor Pyles[3], BA; Janne Marieke Veerbeek[1], MSc, PhD; Felix Bork[4], MSc; Sandro-Michael Heining[5], MD; Nassir Navab[4,6], PhD; Andreas Rüdiger Luft[1,7], MD

[1]Division of Vascular Neurology and Neurorehabilitation, Department of Neurology, University of Zurich and University Hospital Zurich, Zurich, Switzerland

[2]University Hospital Klinikum rechts der Isar, Technical University of Munich, Munich, Germany

[3]Johns Hopkins Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, United States

[4]Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, Munich, Germany

[5]Department of Trauma Surgery, University of Zurich and University Hospital Zurich, Zurich, Switzerland

[6]Computer Aided Medical Procedures, Johns Hopkins University, Baltimore, MD, United States

[7]cereneo Center for Neurology and Rehabilitation, Vitznau, Switzerland

[*]these authors contributed equally

**Corresponding Author:**
Jeremia Philipp Oskar Held, MSc, PhD
Division of Vascular Neurology and Neurorehabilitation
Department of Neurology
University of Zurich and University Hospital Zurich
Frauenklinikstrasse 26
Zurich, 8091
Switzerland
Phone: 41 44 255 5645
Email: jeremia.held@uzh.ch

## Abstract

**Background:** Gait and balance impairments are common in neurological diseases, including stroke, and negatively affect patients' quality of life. Improving balance and gait are among the main goals of rehabilitation. Rehabilitation is mainly performed in clinics, which lack context specificity; therefore, training in the patient's home environment is preferable. In the last decade, developed rehabilitation technologies such as virtual reality and augmented reality (AR) have enabled gait and balance training outside clinics. Here, we propose a new method for gait rehabilitation in persons who have had a stroke in which mobile AR technology and a sensor-based motion capture system are combined to provide fine-grained feedback on gait performance in real time.

**Objective:** The aims of this study were (1) to investigate manipulation of the gait pattern of persons who have had a stroke based on virtual augmentation during overground walking compared to walking without AR performance feedback and (2) to investigate the usability of the AR system.

**Methods:** We developed the ARISE (Augmented Reality for gait Impairments after StrokE) system, in which we combined a development version of HoloLens 2 smart glasses (Microsoft Corporation) with a sensor-based motion capture system. One patient with chronic minor gait impairment poststroke completed clinical gait assessments and an AR parkour course with patient-centered performance gait feedback. The movement kinematics during gait as well as the usability and safety of the system were evaluated.

**Results:** The patient changed his gait pattern during AR parkour compared to the pattern observed during the clinical gait assessments. He recognized the virtual objects and ranked the usability of the ARISE system as excellent. In addition, the patient stated that the system would complement his standard gait therapy. Except for the symptom of exhilaration, no adverse events occurred.

**Conclusions:** This project provided the first evidence of gait adaptation during overground walking based on real-time feedback through visual and auditory augmentation. The system has potential to provide gait and balance rehabilitation outside the clinic. This initial investigation of AR rehabilitation may aid the development and investigation of new gait and balance therapies.

*(JMIR Mhealth Uhealth 2020;8(5):e17804)* doi:10.2196/17804

**KEYWORDS**

HoloLens 2; gait; rehabilitation; stroke; augmented reality; sensors

## Introduction

Many neurological diseases lead to impairments of gait and balance [1]. Approximately 80% of persons who have a stroke experience such deficits [2]. The key characteristics of impaired gait after stroke are shortened stance phase of the affected side, reduced knee and hip flexion during swing, slower walking speed, and shorter stride length [3]. Balance impairments are characterized by reduction in maximum weight shift toward the affected side [4], delayed postural reactions [5], shift of the center of mass toward the non-affected side, and decreased ability to avoid obstacles [6]. These gait and balance impairments are increased when patients are required to perform a cognitive task in parallel [7] and in patients who have attention or vision deficits [8,9]. Impaired gait and balance poststroke can have profound consequences for patients, as these impairments are strongly related to increased fear of falling [10] and reduced quality of life [11].

Therefore, improving gait and balance is one of the main goals of stroke rehabilitation. Repetitive practice is an essential ingredient of established evidence-based interventions [12] such as speed-dependent treadmill training, postural control with visual feedback training, and task-specific training [13]. These interventions can be delivered as part of inpatient or outpatient programs. In all these settings, training lacks context specificity (i.e., mobility in the daily life environment of the individual patient). In the last decade, virtual reality (VR) training systems have been found to effectively improve gait and balance after stroke [14,15]. VR systems provide challenging training situations and many different training environments. However, these VR systems still do not reflect the real-world environment of the patient; also, they require expensive stationary equipment. In addition, these systems are often limited in their variety of exercises to improve balance during gait training [16]. To interact with the real-world environment, augmented reality (AR) is an option to provide multiple sensory feedback enhanced by computer-generated perceptual information.

Considering the importance of performance feedback for motor skill learning [17], AR combined with sensor-based kinematic measurements can deliver fine-grained visual and auditory feedback on gait and balance parameters; this feedback can provide higher specificity and continuity and lower delay than the feedback delivered by a therapist [18]. Sensor-based motion capture systems, including inertial measurement units, can measure a patient's gait kinematics and center of mass outside the laboratory and clinic [19], although laboratory-based optical tracking systems remain the gold standard with respect to the sensitivity and accuracy of these systems [20].

Here, we propose a new method for gait and balance rehabilitation in patients who have had a stroke that has potential to provide challenging and personalized gait and balance therapy with auditory and visual performance feedback based on gait kinematics in an environment that is adjusted to the patients' needs. We developed the Augmented Reality for gait Impairments after StrokE (ARISE) system, in which we combined a development version of a head-mounted system for real-time visual and auditory feedback and a commercially available sensor-based motion capture system. Subsequently, we evaluated the usability of the AR feedback prototype in a chronic stroke subject with minor gait and balance impairments. We hypothesized that this system is more capable of modifying gait kinematics than walking without the system.

## Methods

### Ethics Statement

Ethical clearance to execute the experiment with the presented subject was provided by the Cantonal Ethics Commission of the Canton of Zurich, Switzerland (BASEC-Nr. Req-2019-00758). The participant received information about the experiments. Written informed consent in accord with the Declaration of Helsinki was obtained from the participant prior to performing the experiments.

### Patient Information

The patient was a right-handed man aged 74 years who had a right-hemispheric ischemic stroke in the thalamus, capsula interna, and right temporal lobe 7 years before participating in the experiment. Acutely after stroke, he had mild motor deficits in the left leg (National Institutes of Health Stroke Scale leg item score 1/4). At the time of this experiment, he had minor limitations of sitting and standing balance (Berg Balance Scale 54/56); he also had mild limitations in motor function (Fugl-Meyer Motor Assessment lower extremity subscale 29/34) and strength (Motricity Index lower extremity subscale 88/100) of the affected left side. The patient walked independently without a walking aid (Functional Ambulation Categories 5/5), with a comfortable walking speed of 1.0 m/s and a step length of 0.56 meters as measured by the 10-Meter Walk Test. The patient had slight risk of falling (Dynamic Gait Index 19/24). He reported numbness in the feet and a feeling of wearing socks when he was not. In addition, a mild cognitive impairment was present (Montreal Cognitive Assessment 25/30).

### Materials

The ARISE system consisted of two essential components: an optical see-through head-mounted display (OST-HMD) and a sensor-based motion capture system (Figure 1, Supplementary Material A).

We identified the HoloLens 2 (Microsoft Corporation) as the most suitable OST-HMD, as it provides a wider field of view (43×29 degrees) compared to other devices and thus can display more virtual objects in a real-world environment. In addition, the HoloLens 2 is able to track head movements with an inertial measurement unit and has an intuitive hand-interaction user interface that is enabled through fully articulated hand tracking.

The OST-HMD was used to visualize the AR parkour course (Figure 2 A-C). The state-of-the-art parkour course had an area of approximately 14×4 meters; it consisted of visualizations of real-life obstacles and barriers, such as blocks and floor mats, that forced the patient to perform certain leg movements. A trail of arrows indicated the walking direction, and the parkour course changed dynamically depending on the position of the patient.

**Figure 1.** Patient who has had a stroke wearing the ARISE system, including the optical see-through head-mounted display (HoloLens 2) and the sensor-based motion capture system (Xsens MVN).

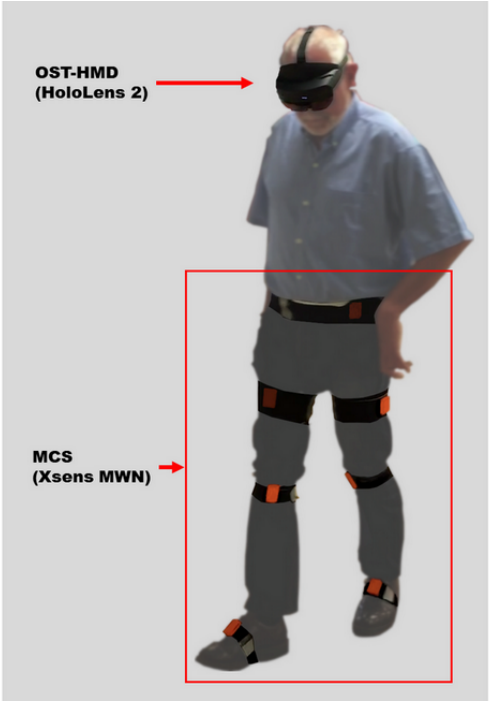
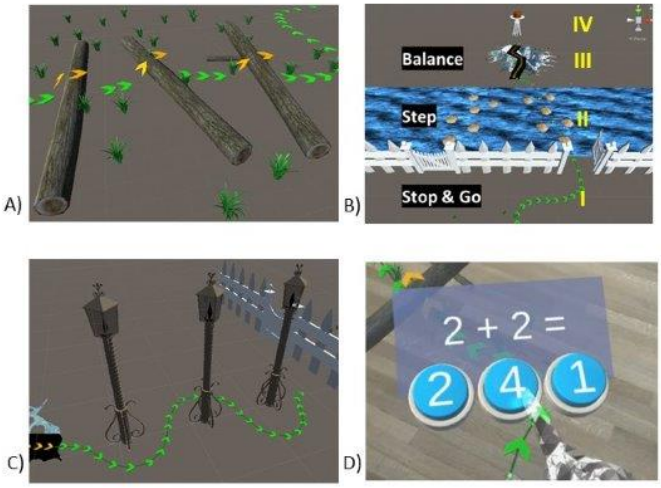



**Figure 2.** Augmented reality parkour course, including arrows indicating the walking direction. A) Overstep obstacle of tree trunks. B) I. Stop and go barrier; II. stepping stones over a virtual river; III. walkable ridge-path; IV. the patient turns around and walks back. C) Walking slalom with lamps. D) Dual-task math calculation.




In addition, the OST-HMD provided visual and auditive feedback based on real-time gait kinematic performance. An inertial measurement unit–based motion capture system, the Xsens MVN (Xsens Technologies B.V.), was chosen to track the kinematics of the lower limbs during gait. We strapped seven inertial measurement units on the pelvis and the lower extremity of the patient (Figure 1). The Xsens software (MVN version 2019.2) converted the rotational data from the inertial

measurement units into a fully articulated virtual mannequin. This provided translational and rotational data of every large humanoid joint. With this combination, the patient was able to walk longer distances (more than 10 meters).

The HoloLens 2 and Xsens systems were integrated by streaming movement performance data to the OST-HMD through a user datagram protocol client. To adjust for translational drift of the motion capture system in the ARISE system, the captured motion data were rigidly attached to the pose of the HoloLens 2. To synchronize the forward directions of both systems, we performed a short calibration phase where the user was required to align his head with his hips and feet. Once the 2 systems were coupled, we were able to provide subtle positive reinforcement of successful knee flexion (>45 degrees) by playing an ambient bird song sound through the HMD.

To increase the difficulty of the tasks the subject was asked to complete, simple math calculations were presented visually (dual task procedure, Figure 2 D). The subject responded by pressing a virtual button, which was detected by the hand-tracking capabilities of the HoloLens 2. At the end of the parkour course, a knowledge of results display was shown, including the time to run the parkour course, the correct answers to the math calculations, and the percentage of time that the condition of knee flexion >45 degrees was fulfilled.

We provided the participant with information about the experimental setup. After clinical assessments of the participant were performed, the sensors were fitted to his legs and pelvis, and he donned the HoloLens 2. First, a patient-personalized height scaling model was applied to the motion capture software. The motion capture system was calibrated. The participant performed a 10-meter walk test and then completed the AR parkour course three times. During walking, the 10-meter walk test and the AR parkour, the patient's lower extremity kinematics and center of mass were tracked with the motion capture system. After the AR parkour, the System Usability Scale [21], the Virtual Reality Symptom Questionnaire [22], and a semi-structured interview were used to assess the patient's experience using the AR system.

### Data Availability

All datasets generated for this publication are included in the manuscript.

## Results

The kinematic parameters of the knees and the center of mass while walking during the 10-meter walk test and the AR parkour are listed in detail in Table 1 and Table 2, respectively. The patient adapted his gait performance and increased his knee flexion angles during the gait cycle when performing the AR parkour compared to when walking a straight line during the 10-meter walk test (Figure 3). His walking speed in the 10-meter walk test did not change (pre 9 seconds, post 9 seconds).

During the AR parkour, the patient perceived the virtual objects, stepped over the obstacles and barriers, and reported a feeling of being in a real-world parkour course. Despite the positive effects on knee flexion, he did not consciously perceive the bird songs.

The patient overlooked the math problems during the first run but solved them in subsequent trials after being reminded to do so. This increased the time to complete the AR parkour from 75 seconds without the second task to 100 seconds with the second task.

The patient reported that wearing the HoloLens 2 felt like wearing a hat. He criticized the limited vertical field of view, which forced him to lean forward to see the obstacles directly beneath him (see the video in Multimedia Appendix 1). However, he ranked the usability of the AR parkour system as excellent (System Usability Scale 87.5/100). He stated that he would like to use the system more often for self-training to complement his current conventional outpatient therapy program. No adverse events were measured with the Virtual Reality Symptom Questionnaire, with the exception of the symptom of exhilaration.

**Table 1.** Kinematic parameters (joint angle, degrees) during the 10-meter walk test and the AR parkour.

| Parameter | Stance[a] | | Swing[b] | | At foot strike[c] | At foot release[d] |
|---|---|---|---|---|---|---|
| | Minimum | Maximum | Minimum | Maximum | Mean (SD) | Mean (SD) |
| **Ten-meter walk test** | | | | | | |
| **Hip angle flexion** | | | | | | |
| Left[e] | –7.17 | 31.06 | –3.20 | 33.57 | 28.34 (2.02) | –0.78 (0.54) |
| Right | –11.18 | 29.63 | –8.01 | 30.21 | 26.74 (0.84) | –4.00 (2.79) |
| Difference | 4.01 | 1.44 | 4.81 | 3.36 | 1.60 | 3.23 |
| **Knee angle flexion** | | | | | | |
| Left[e] | 7.18 | 37.82 | 3.05 | 61.37 | 8.44 (1.59) | 41.84 (0.38) |
| Right | 2.53 | 41.45 | –2.50 | 61.34 | 3.04 (0.73) | 38.23 (4.77) |
| Difference | 4.65 | –3.63 | 5.55 | 0.03 | 5.40 | 3.61 |
| **Ankle angle flexion** | | | | | | |
| Left[e] | –4.92 | 19.49 | –10.85 | 9.07 | 0.86 (1.19) | –4.05 (2.50) |
| Right | –8.11 | 21.58 | –26.09 | 7.63 | –3.14 (2.25) | –4.62 (4.02) |
| Difference | 3.20 | –2.09 | 15.24 | 1.44 | 4.00 | 0.57 |
| **Augmented reality parkour** | | | | | | |
| **Hip angle flexion** | | | | | | |
| Left[e] | –2.09 | 29.77 | –2.06 | 49.26 | 23.77 (3.42) | 3.82 (4.80) |
| Right | –3.55 | 33.90 | –3.30 | 50.98 | 23.49 (4.68) | 1.83 (5.40) |
| Difference | 1.45 | –4.12 | 1.25 | –1.72 | 0.27 | 1.98 |
| **Knee angle flexion** | | | | | | |
| Left[e] | 5.26 | 28.28 | 5.63 | 79.74 | 10.22 (2.84) | 23.06 (4.39) |
| Right | 3.00 | 43.83 | –2.57 | 82.93 | 8.57 (5.42) | 28.08 (6.04) |
| Difference | 2.26 | –15.55 | 8.19 | –3.19 | 1.65 | –5.02 |
| **Ankle angle flexion** | | | | | | |
| Left[e] | –6.70 | 23.22 | –4.49 | 21.33 | 3.60 (3.67) | 15.28 (4.60) |
| Right | –6.32 | 31.45 | –16.95 | 29.63 | 2.13 (3.41) | 17.48 (6.35) |
| Difference | –0.38 | –8.24 | 12.46 | –8.31 | 1.47 | –2.20 |

[a]Period of time from a foot strike to the following release of the same foot.

[b]Period of time from a foot release to the following strike of the same foot.

[c]Time at which the foot contacts the ground after a swing phase.

[d]Time at which the foot stops being in contact with the ground after a stance phase.

[e]The patient's affected side.

**Table 2.** Position of the center of mass in meters during the 10-meter walk test and the augmented reality parkour.

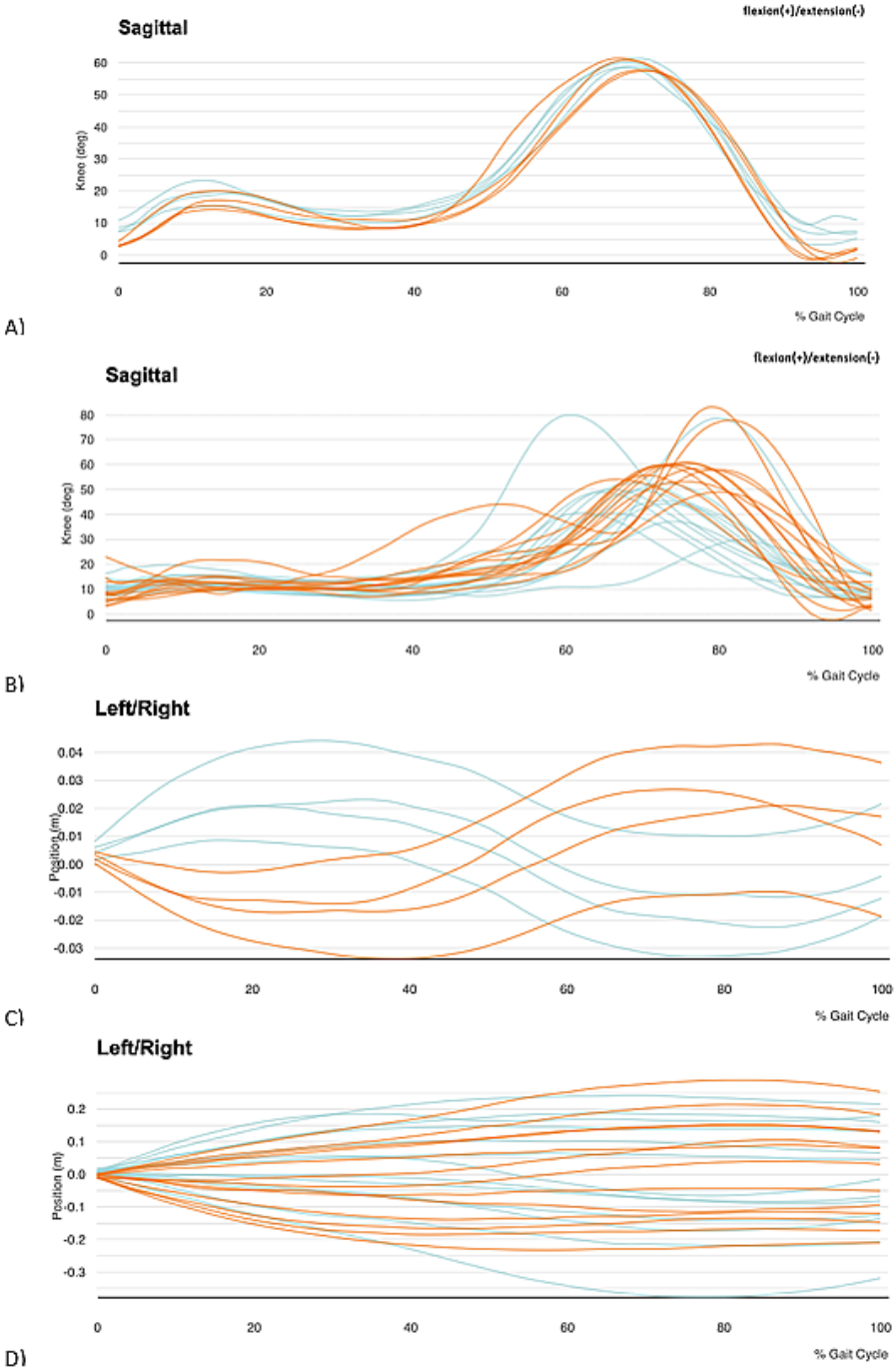
| Position | Stance[a] | | Swing[b] | | Overall | |
|---|---|---|---|---|---|---|
| | Minimum | Maximum | Minimum | Maximum | Minimum | Maximum |
| **Ten-meter walk test** | | | | | | |
| Left[c] | –0.02 | 0.04 | –0.03 | 0.02 | –0.03 | 0.04 |
| Right | –0.03 | 0.02 | –0.02 | 0.04 | –0.03 | 0.04 |
| Difference | 0.01 | 0.02 | –0.01 | –0.02 | 0.00 | 0.00 |
| **Augmented reality parkour** | | | | | | |
| Left[c] | –0.30 | 0.24 | –0.38 | 0.24 | –0.38 | 0.24 |
| Right | –0.23 | 0.26 | –0.23 | 0.29 | –0.23 | 0.29 |
| Difference | –0.07 | –0.02 | –0.15 | 0.05 | –0.15 | –0.05 |

[a]Period of time from a foot strike to the following release of the same foot.

[b]Period of time from a foot release to the following strike of the same foot.

[c]The patient's affected side.

**Figure 3.** Knee flexion angles during gait cycles during A) a 10-meter walking test and B) the AR parkour. Orange lines represent the right knee, and blue lines represent the left knee. Center of mass position during gait cycles during C) the 10-meter walking test and D) the AR parkour. Orange lines represent the right leg, and blue lines represent the left leg.




## Discussion

### Principal Findings

Gait and balance training requires training systems that are mobile and preferably wearable. The ARISE system proposed here combines flexible training with performance feedback in a wearable form by integrating an OST-HMD and a sensor-based motion capture system. It enables context-specific training and can display virtual objects via the HMD in a real-world space. The AR rehabilitation system can be used for overground walking, while gait kinematics are measured with the motion capture system. The combination of these technologies is new in gait rehabilitation. Most AR and VR systems consist of stationary equipment, including a treadmill, cameras, and projection devices [6]. Hence, they offer fewer options for adaptation [23] and are restricted to the unreal training environment of treadmill walking; moreover, a minimum of 6 minutes (around 400 strides) is required to achieve stable performance on a treadmill [24].

The ARISE system was tested during a single session experiment with one patient who had a stroke. The patient perceived the system to be comfortable, and it did not restrict his movements. The limited vertical field of view (29 degrees)

of the HoloLens 2 should be improved to increase the usability of the system. The field of view forced the patient to bend forward; this can induce adverse events, such as near-falls or neck pain, when using the system over longer periods of time.

Through the coupling of the OST-HMD and the motion capture system, the patient perceived auditive feedback based on the kinematics of the knee (flexion angle). The OST-HMD provided feedback in the form of bird songs and a knowledge of results display. When the patient was given multimodal feedback, he changed his gait pattern. This is in line with previous results [6,20]. A transfer of training from the ARISE system to overground walking can be expected when increasing the duration of the intervention [13]. Consolidation may be further improved by adding monetary rewards [25].

The current AR parkour course has a size of 14×4 meters; therefore, it requires a large open space to operate. In future versions, the ARISE system will be able to generate a parkour course with an adaptable area depending on the local environment. This will provide the advantages of flexibility in choosing the test grounds and a variety of parkour designs. In future research, the impacts of different virtual obstacles or even obstacle themes (e.g., street, garden, supermarket) on the gait kinematics can be evaluated. Due to the high mobility of all involved components, our system can be used virtually anywhere, both indoors and outdoors and with different surfaces and distractors. The mobility of the system is also ensured through the web interface of the OST-HMD and the motion capture system, which allows remote access. Therefore, a telerehabilitation approach is conceivable.

The present ARISE system is a prototype, and it requires technical support for setup and calibration. Therefore, a reduced sensor set and simplified calibration procedure would increase its usability and applicability. Furthermore, movement sensors are limited by orientation drift when they are used over a long period of time [26]. In this study, we did not use the system long enough to observe this drift. These issues of technical support and drift currently limit the use of the system in telerehabilitation and should be addressed in future studies [27].

## Conclusions

The ARISE system combines an adjustable and personalized training environment with a feedback and monitoring system for gait and balance rehabilitation. This case study showed that use of the system by patients who have had a stroke and who have mild gait and balance impairments is promising. Compared to existing AR systems, the system utilizes the real physical environment to place virtual obstacles that are immediately and intuitively treated as real-world objects. This system can be used to provide feedback on a variety of gait parameters and to implement personalized (dual-task) gait training environments.

## Acknowledgments

This research was supported by the P & K Pühringer Foundation.

## Authors' Contributions

JH and AL conceived the presented idea. JH designed the study, recruited the patient, conducted the clinical assessments, and analyzed the patient data. KY, CP, and JH developed the methodology and conducted the instrumented experiments during Medical Augmented Reality Summer School 2019 at University Hospital Balgrist, Zurich, Switzerland. JH, KY, and JV drafted the manuscript. All authors revised the manuscript and approved the final version.

## Conflicts of Interest

AL is a scientific advisor for Hocoma AG (Volketswil, Switzerland), which develops rehabilitation technology. The ARISE system, which includes HoloLens 2 (Microsoft) in combination with the Xsens MVN, is a research prototype that was developed during the Augmented Reality Summer School at University Hospital Balgrist, Zurich, Switzerland, and is not commercially available.

Multimedia Appendix 1
Augmented Reality parkour for gait Impairments after StrokE.
[MP4 File (MP4 Video), 31316 KB - mhealth_v8i5e17804_app1.mp4 ]

## References

1. Nonnekes J, Goselink RJM, Růžička E, Fasano A, Nutt JG, Bloem BR. Neurological disorders of gait, balance and posture: a sign-based approach. Nat Rev Neurol 2018 Mar;14(3):183-189. [doi: 10.1038/nrneurol.2017.178] [Medline: 29377011]
2. Duncan PW, Zorowitz R, Bates B, Choi JY, Glasberg JJ, Graham GD, et al. Management of Adult Stroke Rehabilitation Care: a clinical practice guideline. Stroke 2005 Sep;36(9):e100-e143. [doi: 10.1161/01.STR.0000180861.54180.FF] [Medline: 16120836]
3. Perry CJ, Kiriella JB, Hawkins KM, Shanahan CJ, Moore AE, Gage WH. The effects of anterior load carriage on lower limb gait parameters during obstacle clearance. Gait Posture 2010 May;32(1):57-61. [doi: 10.1016/j.gaitpost.2010.03.003] [Medline: 20382021]

4. van Dijk MM, Meyer S, Sandstad S, Wiskerke E, Thuwis R, Vandekerckhove C, et al. A cross-sectional study comparing lateral and diagonal maximum weight shift in people with stroke and healthy controls and the correlation with balance, gait and fear of falling. PLoS One 2017;12(8):e0183020 [FREE Full text] [doi: 10.1371/journal.pone.0183020] [Medline: 28809939]
5. Tasseel-Ponche S, Yelnik AP, Bonan IV. Motor strategies of postural control after hemispheric stroke. Neurophysiol Clin 2015 Nov;45(4-5):327-333. [doi: 10.1016/j.neucli.2015.09.003] [Medline: 26520051]
6. van Ooijen MW, Heeren A, Smulders K, Geurts ACH, Janssen TWJ, Beek PJ, et al. Improved gait adjustments after gait adaptability training are associated with reduced attentional demands in persons with stroke. Exp Brain Res 2015 Mar;233(3):1007-1018. [doi: 10.1007/s00221-014-4175-7] [Medline: 25537466]
7. Bowen A, Wenman R, Mickelborough J, Foster J, Hill E, Tallis R. Dual-task effects of talking while walking on velocity and balance following a stroke. Age Ageing 2001 Jul;30(4):319-323. [doi: 10.1093/ageing/30.4.319] [Medline: 11509310]
8. Bonan IV, Guettard E, Leman MC, Colle FM, Yelnik AP. Subjective visual vertical perception relates to balance in acute stroke. Arch Phys Med Rehabil 2006 May;87(5):642-646. [doi: 10.1016/j.apmr.2006.01.019] [Medline: 16635626]
9. Bonan IV, Hubeaux K, Gellez-Leman MC, Guichard JP, Vicaut E, Yelnik AP. Influence of subjective visual vertical misperception on balance recovery after stroke. J Neurol Neurosurg Psychiatry 2007 Jan;78(1):49-55 [FREE Full text] [doi: 10.1136/jnnp.2006.087791] [Medline: 17012343]
10. Schinkel-Ivy A, Inness EL, Mansfield A. Relationships between fear of falling, balance confidence, and control of balance, gait, and reactive stepping in individuals with sub-acute stroke. Gait Posture 2016 Jan;43:154-159. [doi: 10.1016/j.gaitpost.2015.09.015]
11. Park J, Kim T. The effects of balance and gait function on quality of life of stroke patients. NeuroRehabilitation 2019;44(1):37-41. [doi: 10.3233/NRE-182467] [Medline: 30741699]
12. French MA, Moore MF, Pohlig R, Reisman D. Self-efficacy Mediates the Relationship between Balance/Walking Performance, Activity, and Participation after Stroke. Top Stroke Rehabil 2016 Apr;23(2):77-83 [FREE Full text] [doi: 10.1080/10749357.2015.1110306] [Medline: 26653764]
13. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What is the evidence for physical therapy poststroke? A systematic review and meta-analysis. PLoS One 2014 Feb;9(2):e87987 [FREE Full text] [doi: 10.1371/journal.pone.0087987] [Medline: 24505342]
14. Heeren A, van Ooijen M, Geurts ACH, Day BL, Janssen TWJ, Beek PJ, et al. Step by step: a proof of concept study of C-Mill gait adaptability training in the chronic phase after stroke. J Rehabil Med 2013 Jul;45(7):616-622 [FREE Full text] [doi: 10.2340/16501977-1180] [Medline: 23811818]
15. de RIJM, van DPIGL, Meijer JG. Effect of Virtual Reality Training on Balance and Gait Ability in Patients With Stroke: Systematic Review and Meta-Analysis. Phys Ther 2016 Dec;96(12):1905-1918. [doi: 10.2522/ptj.20160054] [Medline: 27174255]
16. van Duijnhoven HJR, Heeren A, Peters MAM, Veerbeek JM, Kwakkel G, Geurts ACH, et al. Effects of Exercise Therapy on Balance Capacity in Chronic Stroke: Systematic Review and Meta-Analysis. Stroke 2016 Oct;47(10):2603-2610. [doi: 10.1161/STROKEAHA.116.013839] [Medline: 27633021]
17. Veerbeek JM, van Wegen E, van Peppen R, van der Wees PJ, Hendriks E, Rietberg M, et al. What Is the Evidence for Physical Therapy Poststroke? A Systematic Review and Meta-Analysis. PLoS ONE 2014 Feb 4;9(2):e87987. [doi: 10.1371/journal.pone.0087987]
18. Stanton R, Ada L, Dean CM, Preston E. Biofeedback improves activities of the lower limb after stroke: a systematic review. J Physiother 2011;57(3):145-155 [FREE Full text] [doi: 10.1016/S1836-9553(11)70035-2] [Medline: 21843829]
19. Reenalda J, Maartens E, Buurke JH, Gruber AH. Kinematics and shock attenuation during a prolonged run on the athletic track as measured with inertial magnetic measurement units. Gait Posture 2019 Feb;68:155-160. [doi: 10.1016/j.gaitpost.2018.11.020] [Medline: 30481697]
20. Karatsidis A, Richards RE, Konrath JM, van den Noort JC, Schepers HM, Bellusci G, et al. Validation of wearable visual feedback for retraining foot progression angle using inertial sensors and an augmented reality headset. J Neuroeng Rehabil 2018 Aug 15;15(1):78 [FREE Full text] [doi: 10.1186/s12984-018-0419-2] [Medline: 30111337]
21. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. Int J Hum-Comput Int 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
22. Ames SL, Wolffsohn JS, McBrien NA. The development of a symptom questionnaire for assessing virtual reality viewing using a head-mounted display. Optom Vis Sci 2005 Mar;82(3):168-176. [Medline: 15767873]
23. Cavalcanti V, Santana M, Gama A, Correia WF. Corrigendum to “Usability Assessments for Augmented Reality Motor Rehabilitation Solutions: A Systematic Review”. Int J Comput Games Technol 2019 Jan 21;2019:1-1. [doi: 10.1155/2019/9325869]
24. Meyer C, Killeen T, Easthope CS, Curt A, Bolliger M, Linnebank M, et al. Familiarization with treadmill walking: How much is enough? Sci Rep 2019 Mar 26;9(1):5232 [FREE Full text] [doi: 10.1038/s41598-019-41721-0] [Medline: 30914746]
25. Widmer M, Held JP, Wittmann F, Lambercy O, Lutz K, Luft AR. Does motivation matter in upper-limb rehabilitation after stroke? ArmeoSenso-Reward: study protocol for a randomized controlled trial. Trials 2017 Dec 02;18(1):580 [FREE Full text] [doi: 10.1186/s13063-017-2328-2] [Medline: 29197412]

26. Roetenberg D, Luinge H, Slycke P. Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors.: Xsens Motion Technologies BV. Tech Rep; 2009. URL: https://www.researchgate.net/publication/239920367_Xsens_MVN_Full_6DOF_human_motion_tracking_using_miniature_inertial_sensors [accessed 2020-04-20]
27. Held JPO, Klaassen B, Eenhoorn A, van Beijnum BF, Buurke JH, Veltink PH, et al. Inertial Sensor Measurements of Upper-Limb Kinematics in Stroke Patients in Clinic and Home Environment. Front Bioeng Biotechnol 2018 Apr;6:27 [FREE Full text] [doi: 10.3389/fbioe.2018.00027] [Medline: 29707537]

## Abbreviations

**AR:** augmented reality
**ARISE:** Augmented Reality for gait Impairments after StrokE
**OST-HMD:** optical see-through head-mounted display
**VR:** virtual reality

*Edited by G Eysenbach; submitted 14.01.20; peer-reviewed by J Salisbury, M Ozkan; comments to author 03.03.20; revised version received 06.03.20; accepted 22.03.20; published 26.05.20.*

*Please cite as:*
*Held JPO, Yu K, Pyles C, Veerbeek JM, Bork F, Heining SM, Navab N, Luft AR*
*Augmented Reality–Based Rehabilitation of Gait Impairments: Case Report*
*JMIR Mhealth Uhealth 2020;8(5):e17804*
*URL: http://mhealth.jmir.org/2020/5/e17804/*
*doi:10.2196/17804*
*PMID:32452815*

Corrigenda and Addenda

# Correction: Postvaccination Fever Response Rates in Children Derived Using the Fever Coach Mobile App: A Retrospective Observational Study

Sang Hyun Ahn[1*], MD; Jooho Zhiang[2*], BS; Hyery Kim[3], MD, PhD; Seyun Chang[4], BS; Jaewon Shin[4], MD; Myeongchan Kim[4], MD; Yura Lee[5], MD, PhD; Jae-Ho Lee[5,6], MD, PhD; Yu Rang Park[7], PhD

[1]Korea Human Resource Development Institute for Health and Welfare, Cheongju, Republic of Korea

[2]Yonsei University College of Medicine, Seoul, Republic of Korea

[3]Department of Pediatrics, University of Ulsan College of Medicine, Asan Medical Center Children's Hospital, Seoul, Republic of Korea

[4]Mobile Doctor Co, Ltd, Seoul, Republic of Korea

[5]Department of Biomedical Informatics, Asan Medical Center, Seoul, Republic of Korea

[6]Department of Emergency Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea

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

[*]these authors contributed equally

**Corresponding Author:**
Yu Rang Park, PhD
Department of Biomedical Systems Informatics
Yonsei University College of Medicine
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2228 2493
Email: yurangpark@yuhs.ac

**Related Article:**

Correction of: https://mhealth.jmir.org/2019/4/e12223/

***(JMIR Mhealth Uhealth 2020;8(5):e18921)*** doi:10.2196/18921

The authors of "Postvaccination Fever Response Rates in Children Derived Using the Fever Coach Mobile App: A Retrospective Observational Study" (JMIR Mhealth Uhealth 2019;7(4):e12223) identified an error in the Acknowledgements section.

In the first part of the Acknowledgements section, the following phrase:

> *This work was supported by Technology Innovation Program (10060085) funded by the Ministry of Trade, Industry, and Energy (MOTIE, KOREA)*

Has been changed to:

> *This work was supported by Technology Innovation Program (20002289) funded by the Ministry of Trade, Industry, and Energy (MOTIE, KOREA)*

The correction will appear in the online version of the paper on the JMIR website on May 7, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

*Submitted 27.03.20; this is a non–peer-reviewed article;accepted 01.04.20; published 07.05.20.*

*Please cite as:*
*Ahn SH, Zhiang J, Kim H, Chang S, Shin J, Kim M, Lee Y, Lee JH, Park YR*
*Correction: Postvaccination Fever Response Rates in Children Derived Using the Fever Coach Mobile App: A Retrospective Observational Study*
*JMIR Mhealth Uhealth 2020;8(5):e18921*
*URL: https://mhealth.jmir.org/2020/5/e18921*
*doi:10.2196/18921*
*PMID:32379698*

Corrigenda and Addenda

# Correction: Using Goal-Directed Design to Create a Mobile Health App to Improve Patient Compliance With Hypertension Self-Management: Development and Deployment

Huilong Duan[1], PhD; Zheyu Wang[1], BSc; Yumeng Ji[1], MSc; Li Ma[2], MSc; Fang Liu[2], MSc; Mingwei Chi[2], BSc; Ning Deng[1*], PhD; Jiye An[1*], PhD

[1]College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Zhejiang University, Hangzhou, China

[2]General Hospital of Ningxia Medical University, Yinchuan, China

[*]these authors contributed equally

**Corresponding Author:**
Ning Deng, PhD
College of Biomedical Engineering and Instrument Science
Ministry of Education Key Laboratory of Biomedical Engineering
Zhejiang University
38 Zheda Rd, Zhouyiqing Bldg 512
Yuquan Campus
Hangzhou
China
Phone: 86 571 2295 2693
Email: zju.dengning@gmail.com

**Related Article:**

Correction of: https://mhealth.jmir.org/2020/2/e14466/

***(JMIR Mhealth Uhealth 2020;8(5):e18859)*** doi:10.2196/18859

In "Using Goal-Directed Design to Create a Mobile Health App to Improve Patient Compliance With Hypertension Self-Management: Development and Deployment" (JMIR Mhealth Uhealth 2020;8(2):e14466), there was an error which was not identified during the proofreading stage.

The first five sentences and the last word of the Methods subsection in the Abstract were missing in the published paper. The original published Methods subsection of the Abstract was incorrectly presented as:

> *Clustering methods based on questionnaire responses were used to group patients. Qualitative interviews were conducted to identify the needs of different groups. In stage 2, several functional modules were designed to meet the needs of different groups based on the results from stage 1. In stage 3, prototypes of functional modules were designed and implemented as a real app. Stage 4 was the deployment process, in which we conducted a pilot study to investigate patient compliance after using the app. Patient compliance was calculated through the frequency with which they took blood pressure measurements. In addition, qualitative interviews were conducted to learn the underlying reasons for the compliance*

The correct Methods subsection of the Abstract is:

> *The goal-directed design method was applied to guide study design. We divided the study into 4 stages. Stages 1 to 3 comprised the development process. To improve the applicability of the goal-directed design method to chronic disease management, we extracted elements of user models concerned with patient compliance and defined a concrete process for user modeling. In stage 1, personas of hypertensive patients were built using qualitative and quantitative methods. Clustering methods based on questionnaire responses were used to group patients. Qualitative interviews were conducted to identify the needs of different groups. In stage 2, several functional modules were designed to meet the needs of different groups based on the results from stage 1. In stage 3, prototypes of functional modules were designed and implemented as a real app. Stage 4 was the deployment process, in which we conducted a pilot study to investigate patient compliance after using the app. Patient compliance was calculated through the frequency with which they took blood pressure measurements. In addition, qualitative interviews*

*were conducted to learn the underlying reasons for the compliance results.*

The changes made do not affect the findings of the study.

In addition, the original published paper contained an error in author affiliation 1. Affiliation 1 was incorrectly listed as:

*College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Hangzhou, China*

The correct listing for author affiliation 1 is:

*College of Biomedical Engineering and Instrument Science, Ministry of Education Key Laboratory of Biomedical Engineering, Zhejiang University, Hangzhou, China*

The correction will appear in the online version of the paper on the JMIR website on May 28, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

*Submitted 23.03.20; this is a non–peer-reviewed article;accepted 26.03.20; published 28.05.20.*

*Please cite as:*
*Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, Deng N, An J*
*Correction: Using Goal-Directed Design to Create a Mobile Health App to Improve Patient Compliance With Hypertension Self-Management: Development and Deployment*
*JMIR Mhealth Uhealth 2020;8(5):e18859*
*URL: https://mhealth.jmir.org/2020/5/e18859*
*doi:10.2196/18859*
*PMID:32463789*

Original Paper

# Mobile Health Daily Life Monitoring for Parkinson Disease: Development and Validation of Ecological Momentary Assessments

Jeroen Habets[1], MD; Margot Heijmans[1], MSc; Christian Herff[1], PhD; Claudia Simons[2,3], MSc, PhD; Albert FG Leentjens[2], MD, PhD, Prof Dr; Yasin Temel[1], MD, PhD, Prof Dr; Mark Kuijf[4], MD, PhD; Pieter Kubben[1,5], MD, PhD

[1]Department of Neurosurgery, School of Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands

[2]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University, Maastricht, Netherlands

[3]GGzE, Institute for Mental Health Care Eindhoven, Eindhoven, Netherlands

[4]Department of Neurology, Maastricht University Medical Center, Maastricht, Netherlands

[5]Department of Neurosurgery, Radboud University Medical Center, Nijmegen, Netherlands

**Corresponding Author:**
Jeroen Habets, MD
Department of Neurosurgery
School of Mental Health and Neuroscience
Maastricht University
Universiteitssingel 50
Maastricht, 6229 ER
Netherlands
Phone: 31 43 3881348
Email: j.habets@maastrichtuniversity.nl

## Abstract

**Background:** Parkinson disease monitoring is currently transitioning from periodic clinical assessments to continuous daily life monitoring in free-living conditions. Traditional Parkinson disease monitoring methods lack intraday fluctuation detection. Electronic diaries (eDiaries) hold the potential to collect subjective experiences on the severity and burden of motor and nonmotor symptoms in free-living conditions.

**Objective:** This study aimed to develop a Parkinson disease–specific eDiary based on ecological momentary assessments (EMAs) and to explore its validation.

**Methods:** An observational cohort of 20 patients with Parkinson disease used the smartphone-based EMA eDiary for 14 consecutive days without adjusting free-living routines. The eDiary app presented an identical questionnaire consisting of questions regarding affect, context, motor and nonmotor symptoms, and motor performance 7 times daily at semirandomized moments. In addition, patients were asked to complete a morning and an evening questionnaire.

**Results:** Mean affect correlated moderate-to-strong and moderate with motor performance (R=0.38 to 0.75; $P$<.001) and motor symptom (R=0.34 to 0.50; $P$<.001) items, respectively. The motor performance showed a weak-to-moderate negative correlation with motor symptoms (R=−0.31 to −0.48; $P$<.001). Mean group answers given for on-medication conditions vs wearing-off-medication conditions differed significantly ($P$<.05); however, not enough questionnaires were completed for the wearing-off-medication condition to reproduce these findings on individual levels.

**Conclusions:** We presented a Parkinson disease–specific EMA eDiary. Correlations between given answers support the internal validity of the eDiary and underline EMA's potential in free-living Parkinson disease monitoring. Careful patient selection and EMA design adjustment to this targeted population and their fluctuations are necessary to generate robust proof of EMA validation in future work. Combining clinical Parkinson disease knowledge with practical EMA experience is inevitable to design and perform studies, which will lead to the successful integration of eDiaries in free-living Parkinson disease monitoring.

***(JMIR Mhealth Uhealth 2020;8(5):e15628)*** doi:10.2196/15628

**KEYWORDS**

ecological momentary assessment; experience sampling method; electronic diary; Parkinson's disease monitoring

# Introduction

## Background

Parkinson disease is a neurodegenerative disorder that is characterized by bradykinesia, rigidity, and tremor. Many patients develop fluctuations in cardinal motor symptoms, such as bradykinesia, tremor, and postural instability, and levodopa-induced dyskinesia [1,2]. Nonmotor symptoms may also show fluctuations during the day [3,4]. Current gold standards in symptom monitoring, such as the Movement Disorders Society (MDS)–Unified Parkinson Disease Rating Scale and the Parkinson Disease Quality of Life-39, are suboptimal to detect such fluctuations over short periods, as they cover a longer temporal domain and require active observed tasks [5,6]. Monitoring methods that can also detect motor and nonmotor fluctuations over shorter periods in free-living conditions can contribute to applying personalized medicine in Parkinson disease [7,8]. Examples of such new methods are telemonitoring [9] and mobile health (mHealth) apps, often including wearable sensors [10-12]. Patients with neurological conditions are believed to be able to use mobile apps [13]; however, the quality, validation, and usability of the available apps are often low [14]. Nonetheless, there have been promising results of using mHealth monitoring systems for motor and nonmotor symptoms of Parkinson disease during free-living situations [15-17].

Electronic diaries (eDiaries) hold the potential to contribute to these new monitoring methods by collecting valuable information on motor symptoms [18,19] and nonmotor symptoms in free-living conditions [16].

Recently published recommendations on Parkinson disease eDiary development by a specific MDS Task Force and Committee underline the relevance and potential of this approach [20,21]. Ecological momentary assessment (EMA), also referred to as an experience sampling method, is a method that collects subjective experiences at multiple, semirandomized moments during a day. Commonly used in psychiatric and psychological populations, it holds the potential for somatic diseases as well [22]. The scarce literature describing EMA in Parkinson disease reports feasibility in small cohorts of up to 5 patients. Reproduction and further investigation of the usefulness and value of EMA in Parkinson disease are needed [4,16,23].

## Objective

We developed the first specific Parkinson disease eDiary using EMA and set the first steps to validate the EMA method in a broad Parkinson disease cohort.

# Methods

## Study Population

We included 20 patients who were diagnosed with Parkinson disease following the UK Parkinson disease Society Brain Bank Diagnostic Criteria, who were aged between 18 and 80 years, who possessed a smartphone (at least Android 4 or iPhone operating system 8), and who had adequate proficiency in the Dutch language. A Montreal Cognitive Assessment scale score lower than 24 was the only exclusion criterion [24]. Demographic and general disease characteristics were collected, such as sex and age, Parkinson disease duration, levodopa equivalent daily dosage (LEDD), number of daily dopaminergic medication intake moments, presence of intraday motor fluctuations, and recent Hoehn and Yahr scores.

## Ecological Momentary Assessment Study Design

Participants enrolled between August 2018 and March 2019 and participated for 14 consecutive days. EMA questionnaires (referred to as beeps) were presented at seven semirandomized moments a day, one beep within every block of 2 hours between 8 AM and 10 PM. The questionnaire had to be opened within 15 min after notification to prevent procrastination. A separate morning questionnaire was available between 4 AM and 1 PM, and an evening questionnaire was available between 8 PM and 4 AM. Answers on statement questions were given on a 7-point Likert scale. The EMA method was executed via the smartphone app, PsyMate [25]. EMA was combined with the use of three wearable sensors containing accelerometers and gyroscopes. Technical details of the protocol design and feasibility analyses are reported earlier [26]. The study protocol was conducted following the Helsinki guidelines and was approved by the local medical ethical committee of Maastricht University Medical Center+.

## Parkinson Disease–Specific Ecological Momentary Assessment Questionnaire

To the best of our knowledge, no specific EMA questionnaire for Parkinson disease exists. On the basis of a literature search and structured interviews with clinical experts, patients, and caregivers, both focused on parameters that differentiate *good* vs *bad* Parkinson moments, we determined the content of the Parkinson disease EMA questionnaire. We consulted an EMA expert group to phrase the specific questions and design the EMA method.

## Data Preparation

Patients with a completion rate lower than 33% were excluded from analyses [27]. Beeps containing missing values because of unfinished questionnaires or digital data transmission failure were excluded.

To analyze positive and negative affect, we calculated the mean of the items *feeling well*, *feeling cheerful*, and *feeling relaxed* and the mean of the items *feeling down*, *feeling fearful*, and *feeling stressed*, respectively. To analyze general motor function, we calculated the mean of the items *ability to perform current activity*, *ability to walk well*, *ability to talk well*, and *to experience steady mobility*. When we refer to the items *mean positive affect*, *mean negative affect*, or *general motor function* in the paper, we are referring to these calculated mean scores. General motor function in the evening questionnaire was calculated as the mean of the evening questionnaire items *ability to dress*, *ability to eat*, *ability to do household activities*, *ability to do personal care*, and *ability to walk*. The evening questionnaire items *experienced many off periods* and *experienced long off periods* were averaged in an item representing off-moment severity during the day.

To represent the change in an item since the last beep, we calculated differences over time scores. The answer to the previous beep (t−1) was subtracted from the answer of the current beep (t). Two beeps are, on average, 2 hours separated from each other. We did not calculate the difference in scores between the first completed beep a day and the last beep of the previous day.

### Statistical Analysis

We analyzed means, standard deviations, and distributions per item. A skewed distribution of answers of an item to the minimum (1) or the maximum (7) is called a floor or a ceiling effect, respectively. If present, we evaluated whether this floor or ceiling effect could be expected and could be accepted or might be based on an invalid, nonspecific, or nonsensitive question and deserved further evaluation.

To validate whether items measure what they are intended to measure, the correlation between an item and a gold standard that measures the same concept can be assessed. If this expected correlation is present, this means the construct validity of that item is proven [20,28]. As there are no validated assessment scales that assess Parkinson disease symptoms as frequent as our EMA beeps, there is a lack of a gold standard measure. Therefore, we assessed the construct validity by analyzing correlations between items from the same beep that are expected to correlate based on clinical knowledge. To further analyze construct validity, we analyzed correlations between the mean answer over all beeps during 1 day and the answer from the corresponding evening questionnaire. For the latter, we excluded days without the completed evening questionnaire. As the theoretically expected correlation of sleep with other symptoms is ambiguous, we excluded the morning questionnaires from validation analyses.

We compared beep answers given in different medication conditions to explore differences in symptom severity. We merged the two transition conditions, from on-medication to off-medication and vice versa, to differentiate 3 conditions: on-medication condition, off-medication condition, and the transition between on- and off-medication condition.

By calculating correlations between scores of items that are expected to correlate, we analyzed the sensitivity of our EMA questionnaire to measure changes over time. We explored the differences between beep answers given in different medication conditions, on-medication condition, off-medication condition, and transitions between the two. The beeps identified as off-medication condition represent the wearing-off-medication condition because the patients were never fully depleted of dopaminergic medication. In the rest of the paper, we will use the term on-beeps and off-beeps to refer to these medication conditions during a completed beep questionnaire. We performed these comparisons on group and individual levels. The significance of differences between the different medication conditions was calculated using Mann-Whitney U tests. Correlations were calculated using Spearman correlation tests. *P* values were corrected with a Bonferroni correction. All the data preparation and statistical analyses were performed in Python Jupyter Notebook 3 using packages pandas (version 0.24.2), Numpy (version 1.16.4), datetime (version 1.0.0), and Scipy (version 1.3.0).

## Results

### Study Population

We included 4 female and 16 male patients with idiopathic Parkinson disease with a mean age of 63 years (SD 7), a mean disease duration of 8 years (SD 6), and a mean LEDD of 770 mg (SD 394); 6 participants were treated with deep brain stimulation for a mean period of 3.3 years (SD 1.5; Table 1). The mean completion rate was 78% out of 98 continuous beeps (SD 12). No participants were excluded based on a too low completion rate (ie, completion rate <33%) [27].

**Table 1.** Demographics of study population.

| Demographics | Values |
|---|---|
| Gender ratio (female:male) | 4:16 |
| Age (years), mean (SD) | 63 (7) |
| Disease duration (years), mean (SD) | 8 (6) |
| Levodopa equivalent daily dosage (mg), mean (SD) | 770 (394) |
| **Deep brain stimulation** | |
| Patients with deep brain stimulation treatment, n | 6 |
| Duration of deep brain stimulation treatment (years), mean (SD) | 3.3 (1.5) |
| **Hoehn and Yahr scale, n (%)** | |
| 1 | 2 (10) |
| 1.5 | 2 (10) |
| 2 | 7 (35) |
| 2.5 | 3 (10) |
| 3 | 3 (15) |
| 4 | 1 (5) |
| Montreal Cognitive Assessment, mean (SD) | 27.6 (1.5) |

## Parkinson Disease Ecological Momentary Assessment Questionnaire Development

Affect and context items from widely applied EMA questionnaires in psychiatry were added [29]. Parkinson disease–specific items are based on a literature search and structured interviews with clinicians, patients, and caregivers. A detailed description of this literature search and the structured interviews can be found in Multimedia Appendix 1.

Repeated discussions with the *EMA expert group* in our institution (among them CS) gave us the following insights into designing a valid EMA questionnaire for patients with Parkinson disease: (1) do not only assess motor symptoms by direct questions about the specific motor symptom, (2) include assessment of the burden or the influence of the symptoms on the patient's performance/well-being, and (3) include items on context (where/with whom/what) and affect and to have the possibility to correct for varying settings or mood fluctuations. On the basis of the advice of the EMA expert group, we consistently phrased the questions as statements in the "I" perspective and tried to avoid confirming ("I do feel...") and denying ("I do not feel...") statements next to each other [27,30]. Furthermore, when translating clinical terms or items from retrospective questionnaires into EMA items, we aimed to maximize face validity by using everyday language. The final EMA questionnaire is shown in Textbox 1.

**Textbox 1.** Parkinson disease ecological momentary assessment questionnaire content, English translation from the original Dutch version. The beep questionnaire, which is presented seven times during the day, represents the four motor domains as well as affect, cognition, context, and motor performance. The evening questionnaire covers off-moments and motor performance over the day, and the morning questionnaire covers sleep.

**Beep questionnaire (semi-random repeated moments)**

- I feel well
- I feel down
- I feel fearful
- I feel stressed
- I feel sleepy
- I am tired
- I am cheerful
- I am relaxed
- I can concentrate well
- I experience hallucinations
- I am at [home, work, travelling, at family/friend's place, in public]
- I am with [nobody, family, partner, colleagues, friends]
- I am doing [work, resting, household/odd jobs, sports, something else]
- I can do this without hinder
- I am comfortable walking/standing
- I can sit or stand still easily
- I can speak easily
- I can walk easily
- I experience tremor
- I am moving slow
- I experience stiffness
- My muscles are tensioned
- I am uncontrollable moving
- I feel … [1: OFF, 2: ON -> OFF, 3: ON, 4: OFF -> ON]
- I took Parkinson medication since last beep [yes, no, I don't recall]

**Morning questionnaire**

- I slept well
- I woke up often last night
- I feel rested
- It was physically difficult to get up
- It was mentally difficult to get up

**Evening questionnaire**

- I had long OFF periods today
- I had many OFF periods today
- Walking went well today
- (un)dressing went well today
- Eating/ drinking went well today
- Personal care went well today
- Household activities went well today
- I was tired today

### Parkinson Disease Ecological Momentary Assessment Questionnaire Validity

Figure 1 shows means and distributions of all participants per item from the beep questionnaire and the evening questionnaire. Positive affect shows a small ceiling effect, whereas negative affect shows a floor effect. Experiencing hallucinations shows a strong floor effect, with only one participant experiencing hallucinations. Positive formulated items on motor functioning show a small floor effect. Experiencing tremor and dyskinesia shows stronger floor effects than experiencing slowness and stiffness.

Construct validity was assessed by evaluating the presence of expected correlations between items (Figure 2). Mean positive and negative affect showed a strong negative correlation with each other (R=−0.71; *P*<.001; Figure 2). Both positive and negative affect scores showed moderate-to-strong correlations with general motor functioning (R=0.75 and R=−0.53, respectively; *P*<.001). Mean positive and negative affect scores showed weak-to-moderate correlations with different motor symptoms (R=−0.37 to −0.49 and R=0.32 to 0.50, respectively; *P*<.001). General motor functioning showed moderate-to-weak correlations with the motor symptoms tremor, slowness, stiffness, and dyskinesia (R=−0.34, −0.47, −0.44, and −0.51, respectively; *P*<.001).

Beep answers on mean affect scores and general motor functioning from 1 day showed moderate correlations with both the items assessing the amount of experienced off-beeps and the general motor performance from the corresponding evening questionnaires in expected directions (R=−0.43 to 0.69; *P*<.001). Beep answers during the day on slowness, stiffness, tremor, and dyskinesia showed weak-to-moderate correlations with general motor functioning answers from the evening questionnaire (R=−0.24 to 0.44; *P*<.001). These items assessing motor symptoms in the beep questionnaires also showed weak-to-moderate correlations in the expected directions with the item assessing off-beeps in the evening questionnaire (R=0.24 to 0.69; *P*<.001). Although dyskinesia is not a typical symptom during off-beeps, it correlated strongly with off-beeps over the whole day (R=0.69; *P*<.001).

The correlations between difference over time scores were less strong as the absolute answers (see Multimedia Appendix 1 for a correlation heatmap of difference over time scores). All correlations were weak to absent.

**Figure 1.** Distribution plots of answers from beep questionnaires. Mean positive and negative affect showed high and low mean answers, respectively. The ability to perform daily life tasks showed moderate-to-high mean answers, whereas the motor symptom items showed low-to-moderate mean answers. All items were statements and were answered on a 7-point Likert scale, ranging from 1 (not at all) to 7 (very). The white dot represents the median answer, the thick black line represents the IQR, and the thin black lines represent the rest of the distribution, calculated as IQR times 1.5. The width of the shapes correlates with the probability that the patient answered the corresponding value.

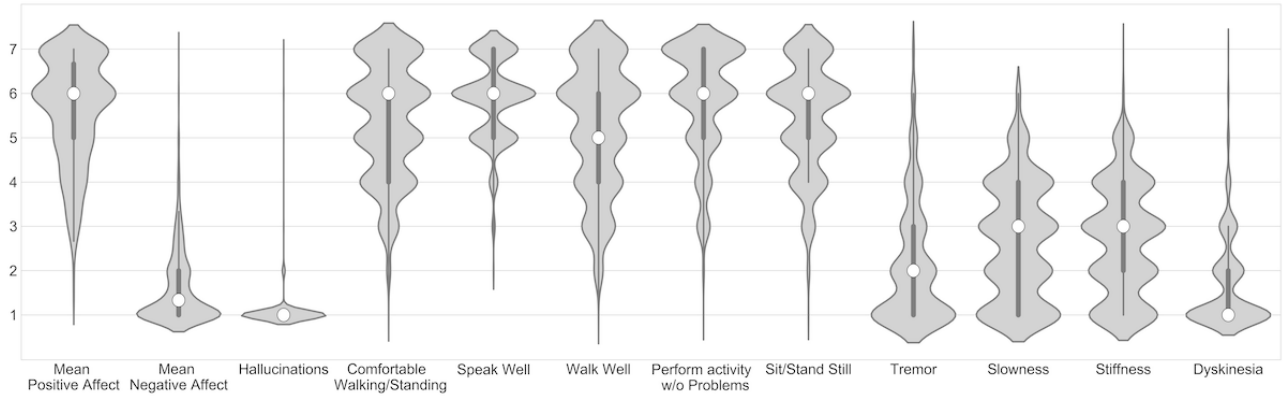
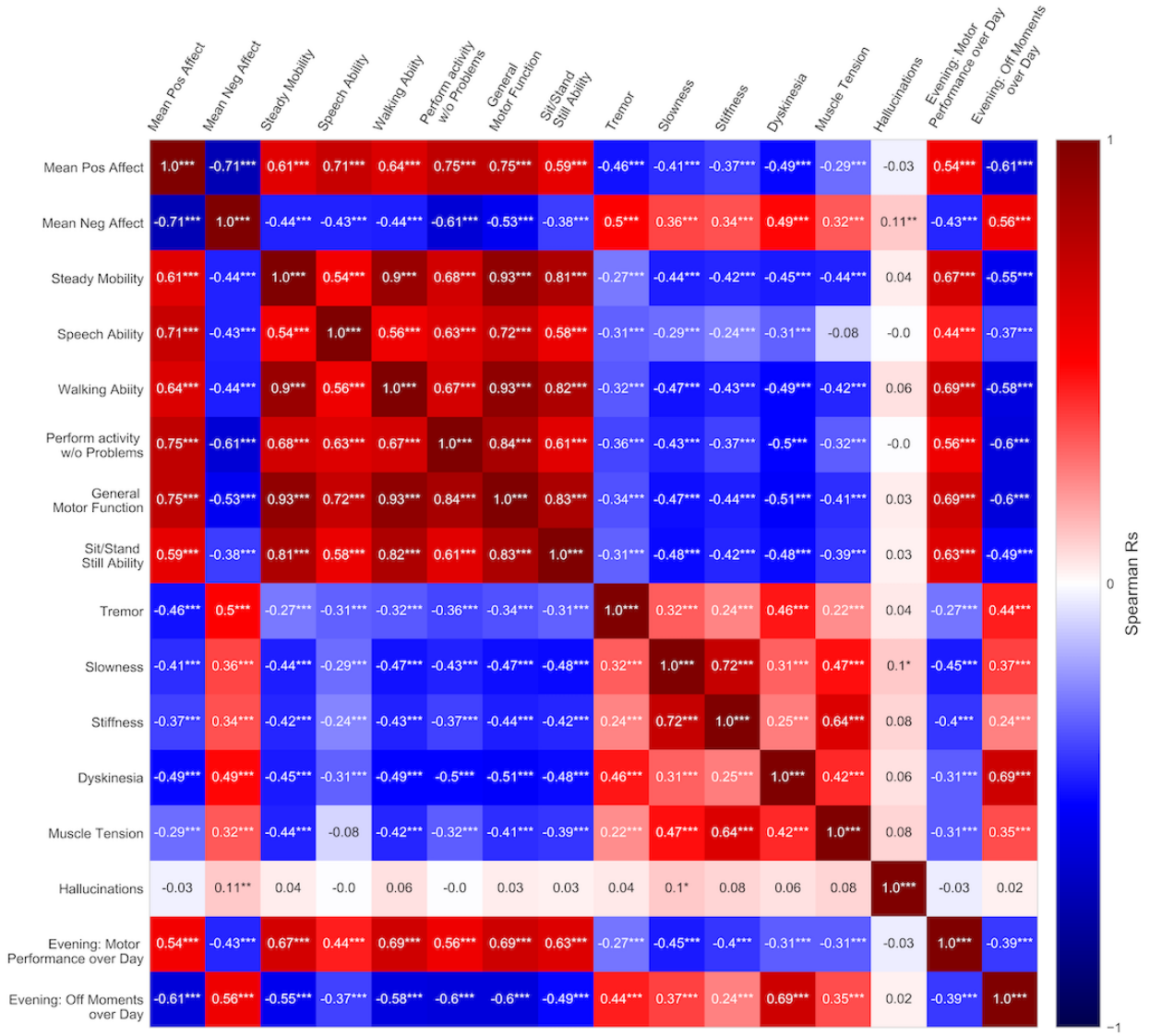


**Figure 2.** Correlations between items from the beep questionnaire and the evening questionnaires. We observed strong and moderate correlations between the motor performance items and mean positive and negative affect, respectively. We observed correlations between mean affect scores, motor symptoms, motor performance, and medication states, in the beep questionnaires and the evening questionnaires, in directions that were expected.

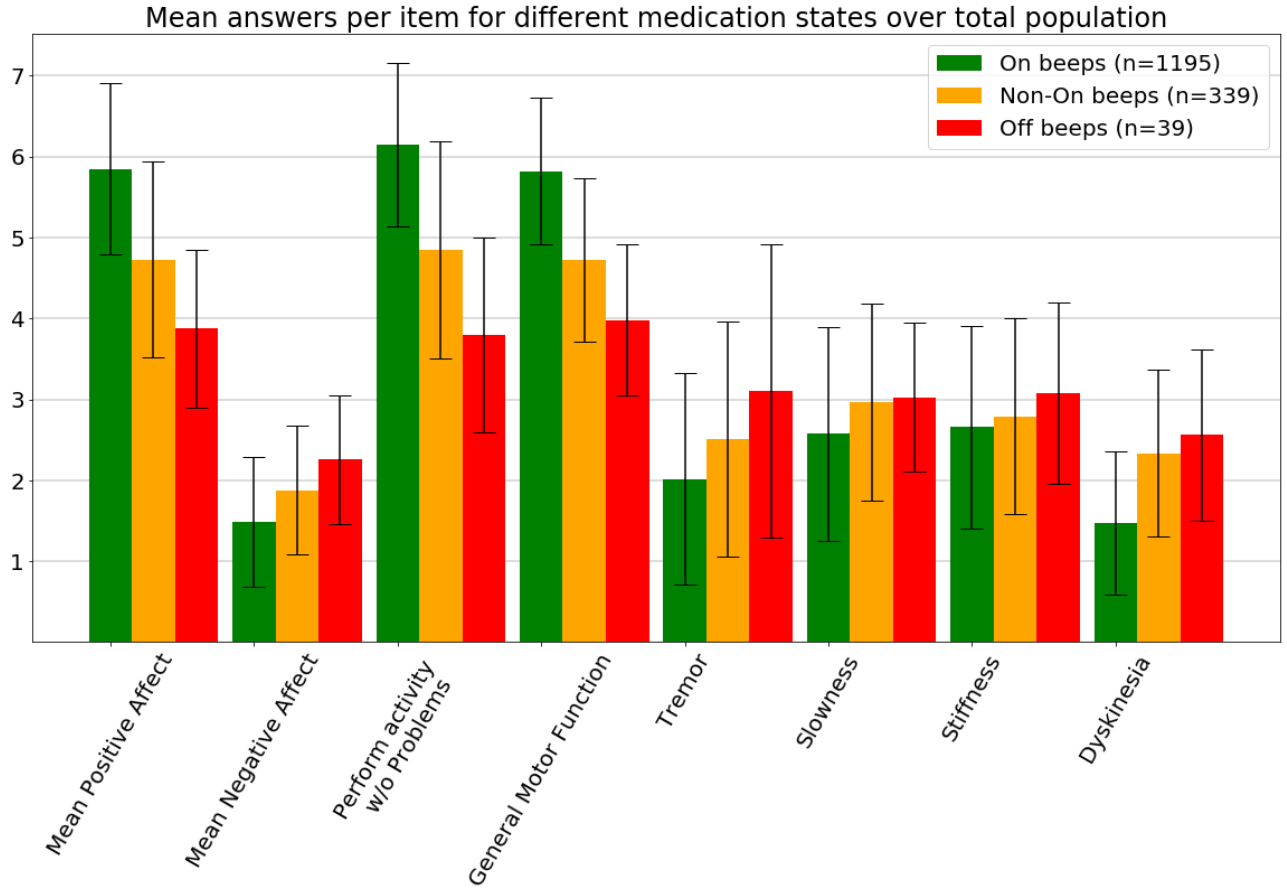



## Influence of Medication Condition on Ecological Momentary Assessment Answers

Of 1573 beeps, 1195 (75.97%) were labeled by patients as *answered in on-medication condition* (on-beeps), 339 (21.56%) were labeled by patients as *answered in between on- and off-medication condition* (transition beeps), and 39 (2.48%) were labeled by patients as *answered in off-medication condition* (off-beeps; Figure 3). On a group level, mean answers significantly differed between on-beeps and non–on-beeps for mean positive affect, general motor function, slowness, and dyskinesia. Mean answers between on-beeps and off-beeps significantly differed for mean positive affect, mean negative affect, general motor function, and tremor.

On an individual level, mean answers during different medication conditions did not differ significantly. The differences were either not significant or not relevant. Only 5 of 20 participants reported 20% or more beeps in the *non–on-medication condition* (Multimedia Appendix 1).

**Figure 3.** Mean answers during different medication states. The given answers in different medication conditions show significant differences. The direction of differences is as expected, except for dyskinesia, which is scored higher, on average, during off-condition compared with on-condition. Whiskers indicate standard deviations. On-beeps represent answers during on-medication, non–on-beeps represent answers during off-medication and during the transition phase, and off-beeps represent answers only during off-medication.




## Discussion

### Clinical Relevance of Ecological Momentary Assessment for Free-Living Parkinson Disease Monitoring

Owing to the fluctuating nature of Parkinson disease and its heterogeneous character, EMA holds theoretically great potential to increase insight into symptom severity and burden fluctuation during free-living conditions. Our work fits in the first milestone defined by the MDS Technology Task Force and the MDS Rating Scales Program Electronic Development Ad-Hoc Committee by giving insight into the prioritization of outcomes, which are relevant for the patient to measure [20]. Obviously, this paper is one of the first of many to follow.

A common challenge for all future work in this field is the validation of methods and questionnaires. Validated scales only exist for Parkinson disease monitoring with longer time intervals than the short time intervals needed to detect intraday fluctuations. This fact makes classical validation with golden standards difficult and even incorrect depending on the methodology. The MDS Task Force on Technology, therefore, advices to validate new Parkinson disease monitoring methods for free-living conditions according to accuracy, reliability, sensitivity, and minimal clinically significant differences [21]. This validation challenge is also relevant for the integration of additional biometric monitor devices, such as accelerometers, gyroscopes, microphones, or electrophysiological monitor devices. Vizcarra et al [20] make a distinction between the integration of action-dependent and action-independent monitoring. It is expected that creating golden standards for action-dependent tasks in, for example, a laboratory setting is easier than creating standards for an action-independent setting such as free-living [18,19]. For the latter, validated Parkinson disease monitoring devices collecting subjective experiences on symptom severity and burden can be of substantial value.

The most applied and promising action-independent Parkinson disease monitoring methods for free-living conditions are based on wearable sensors [31,32]. Attempts to include subjective diary data in the validation of sensor data algorithms were hindered by practical limitations mainly, for example, recall bias and diary fatigue [19,33]. Smartphone-based EMA methodology can be applied less obtrusively and address these traditional diary limitations. Naturally, the feasibility of this method is heavily dependent on the frequency and duration in which it is applied. These factors require thorough future investigations and may differ per intention of use, for example, wearable sensor calibration or periodic free-living monitoring of nonmotor symptoms.

### General Lessons on Ecological Momentary Assessment in Parkinson Disease

The introduction of a new method in Parkinson disease monitoring entails challenges and questions beyond the current literature. To address these challenges and questions as effectively as possible, we gathered a multidisciplinary team consisting of those with clinical Parkinson disease expertise (experts in neurology, neuropsychiatry, and neurosurgery and specialized nurses) and experienced practitioners of EMA (neuropsychology and neuropsychiatry). We described our most important lessons regarding the content and the phrasing of the EMA questionnaire to inform clinicians and researchers interested in applying EMA in Parkinson disease. Moreover, a recently published checklist provides researchers with a tool to design an EMA-based diary study [34]. Essential for EMA in Parkinson disease is the similarity between the frequency of EMA assessments and the frequency of symptom fluctuations that are intended to capture. Thus, EMA studies may require different designs depending on whether they monitor levodopa-induced dyskinesia fluctuations over a day or whether they monitor the effect of an extra levodopa agonist on morning bradykinesia.

### Validation of Ecological Momentary Assessment in Parkinson Disease

Mean answer values and distributions show expected findings (Figure 1). Positive affect items are known to be answered higher than negative affect items [35]. The high mean answers on general motor function and the low mean answers on motor symptoms can be explained by the stable treated population and the relatively low overall disease progression (Table 1). Concerning the observed floor and ceiling effects, we only regard the item on hallucinations as obsolete for this population because of the observed extreme floor effect. As stated earlier, negative affect items are known to show a floor effect. Tremor and dyskinesia also show an unsatisfying floor effect, although we think this is because of the low prevalence of these symptoms in our sample. Moreover, the unexpected positive correlation between dyskinesia and experienced off-beeps suggests that the dyskinesia item might not be well understood by patients. Limited awareness on the presence of dyskinesia among patients with Parkinson disease is described earlier [36]. This finding might also be strengthened by the low prevalence of dyskinesia in the population. We advise, therefore, to avoid the use of nonapplicable, general questions for individual patients. If an item is not applicable for a patient, the patient should be clearly instructed on how to answer that item.

The moderate-to-high correlations present between affect, motor function, and motor symptoms prove the construct validity of the Parkinson disease EMA method partially. The low-to-moderate correlations between motor function and motor symptoms warrant cautious conclusions, and follow-up validation among a narrower selected population with more motor fluctuations is needed to more extensively proof construct validity.

The high number of beeps answered in on-medication condition (Figure 3) and the weak till absent correlations between difference over time scores (Multimedia Appendix 1) confirm this hypothesis. Significance levels are calculated using Mann-Whitney *U* tests (all $P<0.5$). All questions except *Stiffness* differed significantly between on-beeps and non–on-beeps. All questions except *Slowness* and *Stiffness* differed significantly between on-beeps and off-beeps. Ideally, the significant differences that were only found on the group level also hold on individual levels in the next validation study, especially because EMA is intended for individual monitoring.

Despite the fact that further investigation is needed, EMA in Parkinson disease seems to be potentially useful and valid when evaluating the moderate-to-high correlations between affect, general functioning, bradykinesia, and stiffness. Altogether, we interpret our findings as encouraging, and we stress the importance of a careful patient selection depending on the exact goal of EMA monitoring.

### Limitations

The broad inclusion policy was a well-considered choice in the study design, and it resulted in important information about the feasibility and validity of EMA in a broad Parkinson disease population. When applied in a more specified cohort, clinimetric validation analyses necessary for the next step in validation are better feasible, such as principal component analyses to exclude fewer sensitive items. The latter may lead to individual patient- or patient subgroup–specific questionnaire content.

## Conclusions

EMA-based eDiaries are promising to enrich free-living Parkinson disease monitoring with essential information on motor and nonmotor fluctuations. First validation analyses suggest the internal validation of EMA among a general Parkinson disease population. Careful patient selection and EMA design adjustment to this targeted population and their fluctuations are necessary to generate robust proof of EMA validation in future work. Combining clinical Parkinson disease knowledge with practical EMA experience is inevitable to design and perform studies, which will lead to the successful integration of eDiaries in free-living Parkinson disease monitoring.

## Acknowledgments

The authors would like to thank Mirella Waber for advice on the application and content of the eDiary among patients with Parkinson disease, the EMA expert group within the Department of Psychiatry and Neuropsychology at Maastricht University for constructive discussion on the EMA methodology, and Karel Borkelmans for the technical support of the PsyMate app. YT and PK received a grant from Stichting Weijerhorst.

## Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 416 KB - mhealth_v8i5e15628_app1.docx ]

## References


1. Jankovic J. Motor fluctuations and dyskinesias in Parkinson's disease: clinical manifestations. Mov Disord 2005;20(Suppl 11):S11-S16. [doi: 10.1002/mds.20458] [Medline: 15822109]
2. Metman LV. Recognition and treatment of response fluctuations in Parkinson's disease: review article. Amino Acids 2002;23(1-3):141-145. [doi: 10.1007/s00726-001-0119-1] [Medline: 12373528]
3. Kim A, Kim H, Shin CW, Kim A, Kim Y, Jang M, et al. Emergence of non-motor fluctuations with reference to motor fluctuations in Parkinson's disease. Parkinsonism Relat Disord 2018 Sep;54:79-83. [doi: 10.1016/j.parkreldis.2018.04.020] [Medline: 29724602]
4. van der Velden RM, Broen MP, Kuijf ML, Leentjens AF. Frequency of mood and anxiety fluctuations in Parkinson's disease patients with motor fluctuations: A systematic review. Mov Disord 2018 Oct;33(10):1521-1527. [doi: 10.1002/mds.27465] [Medline: 30225905]
5. Goetz CG, Tilley BC, Shaftman SR, Stebbins GT, Fahn S, Martinez-Martin P, Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. Mov Disord 2008 Nov 15;23(15):2129-2170. [doi: 10.1002/mds.22340] [Medline: 19025984]
6. Hagell P, Nygren C. The 39 item Parkinson's disease questionnaire (PDQ-39) revisited: implications for evidence based medicine. J Neurol Neurosurg Psychiatry 2007 Nov;78(11):1191-1198 [FREE Full text] [doi: 10.1136/jnnp.2006.111161] [Medline: 17442762]
7. Sánchez-Ferro Á, Elshehabi M, Godinho C, Salkovic D, Hobert MA, Domingos J, et al. New methods for the assessment of Parkinson's disease (2005 to 2015): A systematic review. Mov Disord 2016 Sep;31(9):1283-1292. [doi: 10.1002/mds.26723] [Medline: 27430969]
8. Dorsey ER, Vlaanderen FP, Engelen LJ, Kieburtz K, Zhu W, Biglan KM, et al. Moving Parkinson care to the home. Mov Disord 2016 Sep;31(9):1258-1262 [FREE Full text] [doi: 10.1002/mds.26744] [Medline: 27501323]
9. Schneider RB, Biglan KM. The promise of telemedicine for chronic neurological disorders: the example of Parkinson's disease. Lancet Neurol 2017 Jul;16(7):541-551. [doi: 10.1016/S1474-4422(17)30167-9] [Medline: 28566190]
10. Cancela J, Mascato SV, Gatsios D, Rigas G, Marcante A, Gentile G, et al. Monitoring of motor and non-motor symptoms of Parkinson's disease through a mHealth platform. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:663-666. [doi: 10.1109/EMBC.2016.7590789] [Medline: 28268415]
11. Tsiouris KM, Gatsios D, Rigas G, Miljkovic D, Seljak BK, Bohanec M, et al. PD_Manager: an mHealth platform for Parkinson's disease patient management. Healthc Technol Lett 2017 Jun;4(3):102-108 [FREE Full text] [doi: 10.1049/htl.2017.0007] [Medline: 28706727]
12. Ferreira JJ, Godinho C, Santos AT, Domingos J, Abreu D, Lobo R, et al. Quantitative home-based assessment of Parkinson's symptoms: the SENSE-PARK feasibility and usability study. BMC Neurol 2015 Jun 10;15:89 [FREE Full text] [doi: 10.1186/s12883-015-0343-z] [Medline: 26059091]
13. Winberg C, Kylberg M, Pettersson C, Harnett T, Hedvall P, Mattsson T, et al. The use of apps for health in persons with multiple sclerosis, Parkinson's disease and stroke - barriers and facilitators. Stud Health Technol Inform 2017;242:638-641. [Medline: 28873864]
14. Rey ML, Vela-Desojo L, Cuerda RC. Mobile phone applications in Parkinson's disease: A systematic review. Neurologia 2019;34(1):38-54 [FREE Full text] [doi: 10.1016/j.nrl.2017.03.006] [Medline: 28549757]
15. Zhan A, Mohan S, Tarolli C, Schneider RB, Adams JL, Sharma S, et al. Using smartphones and machine learning to quantify Parkinson disease severity: the mobile Parkinson disease score. JAMA Neurol 2018 Jul 1;75(7):876-880 [FREE Full text] [doi: 10.1001/jamaneurol.2018.0809] [Medline: 29582075]
16. Broen MP, Marsman VA, Kuijf ML, van Oostenbrugge RJ, van Os J, Leentjens AF. Unraveling the relationship between motor symptoms, affective states and contextual factors in Parkinson's disease: a feasibility study of the experience sampling method. PLoS One 2016;11(3):e0151195 [FREE Full text] [doi: 10.1371/journal.pone.0151195] [Medline: 26962853]
17. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
18. Rodríguez-Molinero A, Pérez-López C, Samà A, de Mingo E, Rodríguez-Martín D, Hernández-Vara J, et al. A kinematic sensor and algorithm to detect motor fluctuations in Parkinson disease: validation study under real conditions of use. JMIR Rehabil Assist Technol 2018 Apr 25;5(1):e8 [FREE Full text] [doi: 10.2196/rehab.8335] [Medline: 29695377]

19. Del Din S, Godfrey A, Mazzà C, Lord S, Rochester L. Free-living monitoring of Parkinson's disease: Lessons from the field. Mov Disord 2016 Sep;31(9):1293-1313. [doi: 10.1002/mds.26718] [Medline: 27452964]
20. Vizcarra JA, Sánchez-Ferro Á, Maetzler W, Marsili L, Zavala L, Lang AE, MDS Technology Task Forcethe MDS Rating Scales Program Electronic Development Ad-Hoc Committee. The Parkinson's disease e-diary: Developing a clinical and research tool for the digital age. Mov Disord 2019 May;34(5):676-681. [doi: 10.1002/mds.27673] [Medline: 30901492]
21. Espay AJ, Hausdorff JM, Sánchez-Ferro Á, Klucken J, Merola A, Bonato P, Movement Disorder Society Task Force on Technology. A roadmap for implementation of patient-centered digital outcome measures in Parkinson's disease obtained using mobile health technologies. Mov Disord 2019 May;34(5):657-663. [doi: 10.1002/mds.27671] [Medline: 30901495]
22. Verhagen SJ, Hasmi L, Drukker M, van Os J, Delespaul PA. Use of the experience sampling method in the context of clinical trials. Evid Based Ment Health 2016 Aug;19(3):86-89 [FREE Full text] [doi: 10.1136/ebmental-2016-102418] [Medline: 27443678]
23. van der Velden RM, Mulders AE, Drukker M, Kuijf ML, Leentjens AF. Network analysis of symptoms in a Parkinson patient using experience sampling data: An n = 1 study. Mov Disord 2018 Dec;33(12):1938-1944. [doi: 10.1002/mds.93] [Medline: 30288791]
24. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
25. PsyMate [NL]. URL: http://www.psymate.eu/ [accessed 2020-03-11]
26. Heijmans M, Habets JG, Herff C, Aarts J, Stevens A, Kuijf ML, et al. Monitoring Parkinson's disease symptoms during daily life: a feasibility study. NPJ Parkinsons Dis 2019;5:21 [FREE Full text] [doi: 10.1038/s41531-019-0093-5] [Medline: 31583270]
27. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
28. Ginty A. Construct validity. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer New York; 2013.
29. Simons CJ, Hartmann JA, Kramer I, Menne-Lothmann C, Höhn P, van Bemmel AL, et al. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. Eur Psychiatry 2015 Nov;30(8):900-906. [doi: 10.1016/j.eurpsy.2015.09.004] [Medline: 26647864]
30. Delespaul AE. Assessing Schizophrenia in Daily Life: The Experience Sampling Method. Maastricht, Netherlands: Maastricht University; 1995.
31. Ramdhani RA, Khojandi A, Shylo O, Kopell BH. Optimizing clinical assessments in Parkinson's disease through the use of wearable sensors and data driven modeling. Front Comput Neurosci 2018;12:72 [FREE Full text] [doi: 10.3389/fncom.2018.00072] [Medline: 30254580]
32. Thorp JE, Adamczyk PG, Ploeg H, Pickett KA. Monitoring motor symptoms during activities of daily living in individuals with Parkinson's disease. Front Neurol 2018;9:1036 [FREE Full text] [doi: 10.3389/fneur.2018.01036] [Medline: 30619024]
33. Papapetropoulos SS. Patient diaries as a clinical endpoint in Parkinson's disease clinical trials. CNS Neurosci Ther 2012 May;18(5):380-387 [FREE Full text] [doi: 10.1111/j.1755-5949.2011.00253.x] [Medline: 22070400]
34. Janssens KA, Bos EH, Rosmalen JG, Wichers MC, Riese H. A qualitative approach to guide choices for designing a diary study. BMC Med Res Methodol 2018 Nov 16;18(1):140 [FREE Full text] [doi: 10.1186/s12874-018-0579-6] [Medline: 30445926]
35. Jacobs N, Menne-Lothmann C, Derom C, Thiery E, van Os J, Wichers M. Deconstructing the familiality of variability in momentary negative and positive affect. Acta Psychiatr Scand 2013 Apr;127(4):318-327. [doi: 10.1111/j.1600-0447.2012.01924.x] [Medline: 22906203]
36. Hoff JI, van Hilten BJ, Roos RA. A review of the assessment of dyskinesias. Mov Disord 1999 Sep;14(5):737-743. [doi: 10.1002/1531-8257(199909)14:5<737::aid-mds1005>3.0.co;2-a] [Medline: 10495034]


## Abbreviations

**eDiary:** electronic diary
**EMA:** ecological momentary assessment
**LEDD:** levodopa equivalent daily dosage
**MDS:** Movement Disorders Society

*Edited by G Eysenbach; submitted 26.07.19; peer-reviewed by A Davoudi, F Lanfranchi; comments to author 11.11.19; revised version received 25.11.19; accepted 15.12.19; published 11.05.20.*

*Please cite as:*
*Habets J, Heijmans M, Herff C, Simons C, Leentjens AFG, Temel Y, Kuijf M, Kubben P*
*Mobile Health Daily Life Monitoring for Parkinson Disease: Development and Validation of Ecological Momentary Assessments*
*JMIR Mhealth Uhealth 2020;8(5):e15628*
*URL: https://mhealth.jmir.org/2020/5/e15628*
*doi:10.2196/15628*
*PMID:32339999*

Original Paper

# Associations Between Parent Self-Reported and Accelerometer-Measured Physical Activity and Sedentary Time in Children: Ecological Momentary Assessment Study

Junia N de Brito[1], MBA, MPH; Katie A Loth[1], LD, RD, MPH, PhD; Allan Tate[1], MPH, PhD; Jerica M Berge[1], CFLE, LMFT, MPH, PhD

Department of Family Medicine and Community Health, University of Minnesota, Minneapolis, MN, United States

**Corresponding Author:**
Junia N de Brito, MBA, MPH
Department of Family Medicine and Community Health
University of Minnesota
717 Delaware Street SE
Suite 400
Minneapolis, MN, 55414
United States
Phone: 1 612 625 0931
Email: nogue013@umn.edu

## Abstract

**Background:** Retrospective self-report questionnaires are the most common method for assessing physical activity (PA) and sedentary behavior (SB) in children when the use of objective assessment methods (eg, accelerometry) is cost prohibitive. However, self-report measures have limitations (eg, recall bias). The use of real-time, mobile ecological momentary assessment (EMA) has been proposed to address these shortcomings. The study findings will provide useful information for researchers interested in using EMA surveys for measuring PA and SB in children, particularly when reported by a parent or caregiver.

**Objective:** This study aimed to examine the associations between the parent's EMA report of their child's PA and SB and accelerometer-measured sedentary time (ST), light-intensity PA (LPA), and moderate-to-vigorous–intensity PA (MVPA) and to examine if these associations differed by day of week, sex, and season.

**Methods:** A total of 140 parent-child dyads (mean child age 6.4 years, SD 0.8; n=66 girls; n=21 African American; n=24 American Indian; n=25 Hispanic/Latino; n=24 Hmong; n=22 Somali; and n=24 white) participated in this study. During an 8-day period, parents reported child PA and SB via multiple daily signal contingent EMA surveys, and children wore a hip-mounted accelerometer to objectively measure ST, LPA, and MVPA. Accelerometer data was matched to the time period occurring before parent EMA-report of child PA and SB. Generalized estimating equations with interaction-term analyses were performed to determine whether the relationship between parent-EMA report of child PA and SB and accelerometer-measured ST and LPA and MVPA outcomes differed by day of the week, sex and season.

**Results:** The parent's EMA report of their child's PA and SB was strongly associated with accelerometer-measured ST, LPA, and MVPA. The parent's EMA report of their child's PA was stronger during the weekend than on weekdays for accelerometer-measured ST ($P \leq .001$) and LPA ($P < .001$). For the parent's EMA report of their child's SB, strong associations were observed with accelerometer-measured ST ($P < .001$), LPA ($P = .005$), and MVPA ($P = .008$). The findings related to sex-interaction terms indicated that the association between the parent-reported child's PA via EMA and the accelerometer-measured MVPA was stronger for boys than girls ($P = .02$). The association between the parent's EMA report of their child's PA and SB and accelerometer-measured ST and PA was similar across seasons in this sample (all $P$ values >.31).

**Conclusions:** When the use of accelerometry-based methods is not feasible and in contexts where the parent is able to spend more proximate time observing the child's PA and SB, the parent's EMA report might be a superior method for measuring PA and SB in young children relative to self-report, given the EMA's strong associations with accelerometer-measured PA and ST.

***(JMIR Mhealth Uhealth 2020;8(5):e15458)*** doi:10.2196/15458

**KEYWORDS**

ecological momentary assessment; accelerometry; mobile devices; physical activity; sedentary behavior; children

## Introduction

Reduced physical activity (PA) and increased sedentary behavior (SB) among children have been associated with less healthy body composition, reduced cardiovascular and musculoskeletal fitness, and other health problems [1]. Empirical evidence supports that important ethnic and racial disparities exist with regard to children's engagement in PA and health outcomes associated with low levels of PA [2,3]. Owing to the importance of PA to the health of American children, the second edition of the 2008 Physical Activity Guidelines for Americans recommended that preschool-aged children (ie, aged 3-5 years) engage in active play throughout the day, and children aged 6 to 17 years accumulate at least 60 min of moderate-to-vigorous–intensity PA (MVPA) per day for disease prevention and health promotion [4]. Given these recommendations, it is of paramount importance that continued efforts are made to improve our ability to accurately assess PA and SB in children.

Several techniques have been used to assess SB and/or PA in children, including, but not limited to, doubly labeled water, accelerometry, and retrospective self-report, which have all been validated with children. Doubly labeled water is the gold standard for measuring total energy expenditure and estimating PA level; however, objective assessment of PA using this method is often cost prohibitive in large-scale epidemiologic studies. As an alternate approach, accelerometers have gained popularity as an objective measurement tool because of their feasibility of use in real-world settings and thus have become the method of choice in epidemiological studies and trials [5]. In addition, the use of accelerometers overcomes many of the recall-based limitations of retrospective assessments and provides accurate measurements of both PA and SB [6,7]. Importantly, there is evidence that the use of accelerometers is feasible with young children [8], irrespective of the unpredictable and irregular nature of play behavior [9]. However, both doubly labeled water and accelerometry techniques require considerable expertise. In addition, these methods can be costly and suffer from administrative time burden on the research team [10-12]. Retrospective self-report questionnaires are the most common method for PA and SB assessment in children and can vary by what they measure (eg, intensity, duration, and frequency of PA) [10]. Although these questionnaires are cost-effective, easy to administer, and provide good assessment of discrete categories of activity level (eg, low, moderate, and high), these questionnaires may suffer from measurement errors (ie, overestimation or underestimation of self-reported PA and SB) and rely heavily on participant's recall ability, which may result in recall bias [13,14]. In fact, studies with children have shown that social desirability bias generally attenuates the strength of associations between reports of PA and SB with health outcomes [14,15]. In addition, previous literature showed that objectively measured PA and self-reported PA and/or PA recall assessment for children have low-to-moderate levels of agreement [16,17]. Therefore, technological advances in survey assessment of SB and PA are needed to overcome these common reporting biases. This is of particular relevance for studying children populations because young children are unlikely to provide reliable estimates of time engaged in PA [10]. Thus, researchers must rely on proxy questionnaire reports from parents and/or another adult caregiver, particularly for young children [18].

An innovative technological advance that has been introduced to minimize the limitations associated with retrospective self-reports from children is ecological momentary assessment (EMA). This technique uses real-time assessment of behavior and was developed to combine both ecological and momentary aspects of the behavior being investigated [19]. EMA is particularly suitable for studying PA and SB in children because children's behavior can be sporadic and contextual [20]. Furthermore, studies investigating the child's PA and SB via EMA have used different software apps and/or Web-based tools that can be loaded or prompted in a mobile device for repeated assessments throughout the day. A few studies have investigated the associations between self-reported PA and SB using EMA via mobile devices and accelerometer-measured PA and ST in children [21-23]. Specifically, Zink et al [23] reported that the use of EMA surveys by children is highly correlated with accelerometer measures, and other studies have reported EMA to be a feasible measure of PA and SB with children [21,22]. However, to our knowledge, EMA has not been investigated as a measure of a child's PA and SB when reported by a child's parent or other adult caregiver (hereafter *parent*). In addition, few studies using EMA have focused on populations from diverse ethnic and/or racial backgrounds and low socioeconomic status, which is important, given the health disparities in PA and SB in these populations [3]. Therefore, investigating the use of EMA for parental report of the child's PA and SB and how well it is correlated with accelerometer-measured sedentary time (ST), light PA (LPA), and MVPA in a diverse sample is important and might be a useful option to not only overcome issues related to retrospective self-report measures by children but also for when the use of objective measurement techniques by researchers are cost prohibitive.

The purpose of this study was to investigate the associations between the parent's report of their child's PA and SB via electronically delivered EMA surveys and accelerometer-measured ST, LPA, and MVPA in a diverse sample and if these associations differed by day of week (ie, weekday vs weekend day), sex (ie, boys vs girls), and season (ie, summer vs school year), given that previous studies reported that children's engagement levels in PA and SB are likely to differ by these factors [23-26]. The findings from this study might provide useful information for researchers interested in using the parent-reported, electronically delivered EMA surveys for measuring PA and SB in children.

## Methods

### Data Source

Data used in this study were drawn from phase 1 of the *Family Matters* study. The full rationale and methodology of the *Family Matters* study have been published elsewhere [27]. Briefly, phase I included in-depth mixed methods cross-sectional examinations of the home environment of racially/ethnically diverse children from low-income families (N=150). Inclusion

criteria for phase 1 of the *Family Matters* study included children aged 5 to 7 years with no medical problem precluding study participation (eg, serious mental illness and disease altering diet or PA); a recent and verified medical record of BMI >5th percentile for age and sex; fluency in English, Spanish, Hmong, and/or Somali; reside with the parent participating in the study; and parent and child not currently participating in a weight management program. Participants engaged in a 10-day in-home observation period that included 2 in-home visits and an 8-day observational period in between home visits. The first home visit (baseline) included, but was not limited to, consenting procedures, demographics, anthropometric measurements (ie, weight was measured 2-3 times using a portable digital scale [Seca 869 model] and recorded to the nearest 0.1 kg; height was measured in duplicate using a portable stadiometer [Seca 217 model] and recorded to the nearest 1.0 cm), and EMA and accelerometry training. The present analyses used selected demographic and anthropometric variables and child accelerometry and EMA data from phase 1. The University of Minnesota Institutional Review Board approved phase 1 of the *Family Matters* study, and all participants consented in accordance with the Declaration of Helsinki procedures [28].

## Participants

The analytic sample of this study included 140 parent-child dyads, with children aged 5 to 7 years from 6 racial and ethnic groups (African American, n=21; American Indian, n=24; Hispanic/Latino, n=25; Hmong, n=24; Somali, n=22; and white, n=24). Ten children were excluded because they did not meet the *Family Matters* study protocol for daily accelerometer wear time (ie, 8 hours for 3 weekdays and 1 weekend day). The demographic and anthropometric characteristics of these participants were similar to the full sample population (N=150). On average, the children included in this sample were aged 6.4 (SD 0.8) years, 47.1% were girls (66/140), and 47.8% (67/140) were overweight or obese (BMI percentile ≥85). Parent-child dyads were, on average, primarily low income (98/140, 70.0% with household income <US $34,000) and comprised of young adult parents (129/140, 92.1% female, 11/140, 7.9% male; mean age 35 years, SD 7.1) who worked full time (62/140, 44.3%) and had at least a high-school education (105/140, 75.0%).

## Measures

### *Accelerometer*

Children's PA levels and ST were assessed objectively using a duo-dimensional accelerometer (ActiGraph GT1M; ActiGraph, LLC), which has been previously validated with children [6,24]. The GT1M devices sampled activity in 15-second epochs, which were then processed to determine ST, LPA, and MVPA duration, as determined by Evenson et al [29]. ST corresponded to 0 to 100 counts per min, LPA corresponded to 101 to 2295 counts per min, and MVPA corresponded to 2296 or more counts per min. Accelerometer nonwear time criteria were defined as more than 60 consecutive min of zero counts, and nonwear and nonvalid data were removed before analysis. For this study, a minimum wear time criterion was determined to be 4 days (ie, 3 weekdays and 1 weekend day) over the 8-day observation period, with at least four waking hours of wear time per day. Participants were instructed to wear the accelerometer for all waking hours during 8 consecutive days and to only remove it for sleeping, bathing, and/or water-based activities (eg, swimming). Accelerometers were attached to an elastic belt and fitted on the child's right hip, with the parent's supervision to ensure proper placement. Detailed written and verbal instructions regarding accelerometer wear were provided for both children and parents to facilitate compliance. Families were compensated with a US $25 gift card, in addition to the study incentive of approximately US $300 for complying with the full accelerometry wear-time demands.

## Ecological Momentary Assessment

The parent's report of their child's PA and SB was measured via multiple daily signal-contingent EMA surveys [30] over an 8-day observation period. More specifically, a minimum of two of 4 signal-contingent daily surveys per day was necessary to be considered a valid response day. Parents were provided an iPad (Apple Inc) and received verbal and written instructions during the first in-home visit, in addition to hands-on training on how to use this mobile device to respond to EMA surveys. Signal-contingent EMA recordings were prompted with a beep on the iPad or via text message as a start signal and were programmed to be delivered randomly, 4 times a day, within an interval of 3-hour time block, from 7 to 10 am, 12 to 2 pm, 3 to 6 pm, and 7 to 10 pm. The EMA survey link expired after 1 hour. In addition, the EMA surveys were delivered in the parents' preferred language (ie, English, Spanish, Somali, or Hmong), and the scheduled EMA prompt delivery was adjusted for parent shift work and wake times to accommodate parents' differing life circumstances.

In this study, the following EMA questions were used to examine the parent's report of their child's PA and SB: (1) *Since the last survey/Since you woke up this morning has [child's name] done something physically active?* (yes/no) and (2) *Since the last survey/Since you woke up this morning has [child's name] watched television/movies or played video games*? (yes/no), respectively. The former was used as a proxy measure for the parent's understanding of child engagement in LPA and MVPA, and the latter was used as a proxy measure for the parent's understanding of child engagement in SB. To facilitate compliance, participants were asked to carry the iPad with them throughout the day. Parents were not required to be in the presence of their child when answering signal-contingent EMA survey questions. Other requirements and details regarding the EMA surveys used in the *Family Matters* study have been published elsewhere [27].

## Statistical Analysis

The child's PA and ST were measured at the hour level using an accelerometer. These data were matched to the period occurring before the parent's EMA assessment of their child's PA and SB on the same observation day and period of accelerometer wear time. For example, if an EMA survey was completed at any time during the 10 am hour of a Monday, accelerometer time was matched to that EMA survey report through 9:59 am of that same Monday to ensure the retrospective assessment did not include accelerometer time measured after the parent EMA survey report. Cross-tabulations and descriptive

statistics were used to identify patterns of wear time, ST, LPA, and MVPA over the course of the day. All outcome variables were continuous (ie, minutes of ST, LPA, and MVPA) and were standardized to accelerometer minutes per hour measured in the signal-contingent EMA survey pre-period. The primary predictor variable was dummy-coded parent's report of the child's engagement in SB and PA before the EMA survey assessment.

The generalized estimating equation approach was used as the primary analytical method. Huber/White robust standard errors were used to deal with model misspecification and correlated participant error terms because of repeated measures [31]. A Gaussian variance family with identity link was used, and a within-participant correlation structure was set to independent to preserve statistical precision [32]. Residual versus fitted plots were used to examine model fit, and any patterns in the error term by predicted mean level. Adjusted models included child race/ethnicity, normal weight/overweight status, day of the week, sex, age, season, and household income. Interaction-term analyses were also performed to determine if the relationships between the parent's EMA report of their child's PA and SB and accelerometer-measured ST, LPA, and MVPA outcomes were modified by day of the week (ie, weekday vs weekend day), sex (ie, boys vs girls), and season (ie, summer vs school year). The interaction terms between the parent's EMA report of their child's PA or SB with day of the week, sex, and season were included separately in each model. Adjusted Wald chi-square tests assessed if EMA compliance and accelerometry wear time differed by day of week, child sex, and season. Sensitivity analyses were performed to determine if more strict accelerometer wear time inclusion criteria (≥8 hours) affected the direction and magnitude of association of our findings, and the results were not affected. The 4-hour minimum inclusion criterion was retained for all models. All data management and analyses were performed in Stata 15.1 MP (StataCorp).

# Results

## Ecological Momentary Assessment and Accelerometry Compliance

Of 150 parent-child dyads, 140 complied with the minimum EMA and accelerometry requirements. In total, 3127 EMA surveys were completed and successfully matched to 944 days of eligible accelerometer data. Overall, compliance (ie, minimum of two surveys completed out of total number of surveys completed) of daily signal-contingent EMA surveys was 82.8% (3127/3776), and 88.5% (124/140) of respondents completed at least three or more surveys, indicating high EMA survey engagement. In addition, in the analytic sample, the average child accelerometer wear time was 7.8 hours per day. Neither EMA compliance nor accelerometry wear time differed by day of week, child sex, or season.

## During What Periods Are Children More or Less Physically Active?

Table 1 descriptively shows average hours of child accelerometer wear time as well as minutes per hour of ST, LPA, and MVPA by order and time of EMA survey completion. On average, children wore the accelerometer for approximately 2 hours before the time in which parents answered the EMA survey, with 71% (5.5/7.8 hours) of the wear time encompassing the first 2 daily EMA surveys. Before the first survey, children spent, on average, two-thirds of an hour engaged in ST (38 min per hour), about one-third of an hour in LPA (19 min per hour), and 3 min per hour in MVPA. As the day progressed, less time was spent in ST, and more time was spent in LPA, with modest increases in MVPA. Notably, in between the third and fourth EMA survey answered, children spent half of their time engaged in ST (30 min per hour), 25 min per hour in LPA, and 5 min per hour in MVPA.

**Table 1.** Accelerometer-measured average minutes of the child's physical activity and the composition of physical activity category by the time of ecological momentary assessment survey completion (N=140 for respondents; N=944 for observation days; and N=3127 for ecological momentary assessment surveys).

| Daily order of EMA[a] survey | Average time of EMA survey completion | Accelerometer physical activity category measured before EMA survey completion | | | |
|---|---|---|---|---|---|
| | | Average child wear time (hours) | Sedentary time, minutes/hour (wear time), n (%) | Light-intensity physical activity, minutes per hour (wear time), n (%) | Moderate-to-vigorous–intensity physical activity, minutes per hour (wear time), n (%) |
| First survey | 9:56 AM | 2.1 | 38 (64) | 19 (32) | 3 (4) |
| Second survey | 2:04 PM | 3.4 | 32 (53) | 24 (40) | 4 (7) |
| Third survey | 4:02 PM | 1.7 | 30 (50) | 25 (42) | 5 (8) |
| Fourth survey | 5:27 PM | 0.6 | 30 (5) | 25 (42) | 5 (8) |

[a]EMA: ecological momentary assessment.

## Did Day of the Week Modify the Association Between Parent Ecological Momentary Assessment Report of the Child's Physical Activity and Sedentary Time and the Accelerometer-Measured Sedentary Time, Light-Intensity Physical Activity, and Moderate-to-Vigorous–Intensity Physical Activity?

The average stratum estimates of ST, LPA, and MVPA for day of the week statistical interaction analyses when the parent reported/did not report the child's PA and/or SB are presented in the Multimedia Appendix 1. The results indicated that on

weekends relative to weekdays, the parent's report of their child's PA via EMA was more strongly, negatively associated with accelerometer-measured ST ($P_{interaction}$≤.001) and positively associated with LPA ($P_{interaction}$<.001), but with similar levels of MVPA ($P_{interaction}$=.37) during weekends and weekdays. Notably, on weekends, the parent's EMA report of their child's PA was associated with an average decrease of −5.1 (95% CI −6.9 to −3.3) min per hour of ST and an average increase of 4 (95% CI 2.4 to 5.5) min per hour in LPA and 1.1 (95% CI 0.3 to 1.9) min per hour in MVPA, relative to the parent's EMA report of their child not engaging in PA. The results also indicated that on weekends, the parent's EMA report of their child's SB was strongly positively associated with accelerometer-measured ST ($P_{interaction}$<.001) and negatively associated with both LPA and MVPA ($P_{interaction}$=.005 and .008, respectively). The relationship between the parent's EMA report of their child's SB and accelerometer-measured ST and LPA was stronger on weekends than on weekdays. On weekends, the parent's EMA report of their child's SB was associated with average increase of 2.8 (95% CI 0.9 to 4.7) min per hour of ST and average decrease of −1.6 (95% CI −3.23 to 0.03) min per hour of LPA and −1.2 (95% CI −2.1 to −0.4) min per hour of MVPA relative to the parent's EMA report of their child not engaging in SB.

### Did Sex Modify the Association Between Parent Ecological Momentary Assessment Report of the Child's Physical Activity and Sedentary Behavior and the Accelerometer-Measured Sedentary Time, Light-Intensity Physical Activity, and Moderate-to-Vigorous–Intensity Physical Activity?

The results indicated that the association between the parent's report of their child's PA and accelerometer-measured MVPA differed between boys and girls ($P_{interaction}$=.02) in this sample (see Multimedia Appendix 2).

Specifically, for boys, the parent's EMA report of their child's PA was associated with an average increase of 1.2 (95% CI 0.7 to 1.7) min of MVPA per hour compared with when the parent reported the child was not engaging in PA. For girls, our findings revealed similar amounts of average minutes per hour of MVPA (0.2; 95% CI −0.5 to 0.9) when parents reported the child's PA relative to when they did not report the child's PA via EMA. The statistical interaction analyses for the parent's EMA report of their child's PA and accelerometer-measured ST and LPA association revealed similar results between boys and girls ($P_{interaction}$=.17 and .55, respectively). In addition, our results indicated that the association between the parent's EMA report of their child's SB and accelerometer-measured ST and LPA somewhat differed between boys and girls ($P_{interaction}$=.049 and .05, respectively) in this sample. For boys, the results showed that when the parent reported the child's SB via EMA, it was associated with an average increase of 1.4 (95% CI −0.2 to 3.0) min per hour of ST and an average decrease of −0.9 (95% CI −2.2 to 0.4) min per hour of LPA, relative to when the parent reported the child not engaged in SB. For girls, however, our results revealed that when the parent reported the child engaged in SB via EMA, it was associated with an average decrease of −1.1 (95% CI −3.0 to 0.8) min per hour of ST and an average increase of 1.2 (95% CI −0.4 to 2.9) min per hour of LPA, relative to when the parent reported the child not engaged in SB. According to our statistical interaction analysis, when the parent reported the child engaged in SB, boys and girls had similar levels of MVPA ($P_{interaction}$=.39).

### Did Season Modify the Association Between Parent Ecological Momentary Assessment Report of the Child's Physical Activity and Sedentary Behavior and the Accelerometer-Measured Sedentary Time, Light-Intensity Physical Activity, and Moderate-to-Vigorous–Intensity Physical Activity?

The statistical interaction analysis did not indicate that the association between the parent's EMA report of their child's PA and accelerometer-measured ST, LPA, and MVPA differed by season (ST: $P_{interaction}$=.41; LPA: $P_{interaction}$=.43; and MVPA: $P_{interaction}$=.59) in this sample. Similarly, regarding the association between the parent's EMA report of their child's SB and accelerometer-measured ST, LPA, and MVPA, the results of the interaction analysis revealed similar results across seasons (ST: $P_{interaction}$=.85; LPA: $P_{interaction}$=.89; and MVPA: $P_{interaction}$=.31).

## Discussion

### Principal Findings

This study examined the relationships between the parent's report of their child's PA and SB via electronically delivered EMA surveys and simultaneous objective (ie, accelerometry) measurement of child engagement in minutes per hour of ST and PA (ie, LPA and MVPA) and if these associations differed by day of the week, sex, and season. The results revealed that the use of mobile EMA surveys by parents to report the child's PA and SB was strongly associated with accelerometer-measured ST, LPA, and MVPA in this sample. Day of the week and sex were identified as moderators of these relationships. These findings suggest that the parent-reported EMA surveys might be a suitable measure for capturing PA and SB in children, particularly when the parent is able to spend more proximate time observing the child's sedentary and PA behaviors.

Our results showed a stronger association between the parent's EMA report of their child's PA and SB and accelerometer-measured ST, LPA, and MVPA and ST and LPA, respectively, for weekend days than on weekdays. These results are consistent with other studies [24,26,33]. Given that the majority of our sample included parents who worked full time, it is plausible that parents were more often in the presence of their child during weekends and, therefore, were able to provide better estimates of their child's PA and SB during weekend days relative to weekdays via EMA. Previous research indicated that retrospective, parent self-reported PA and SB methods are not a suitable proxy measure for assessing 2- to 9-year-old children's PA and SB [34]. In fact, these retrospective methods rely on the respondent's memory, which can lead to recall bias [35]. Given that mobile technology is now widely available,

researchers may rely on electronically delivered EMA surveys to overcome these self-report issues related to more traditional retrospective assessments (eg, 7-day physical activity recall). Therefore, EMA might be advantageous over retrospective survey methods because EMA surveys can be programmed to be delivered multiple times during the day, with such prompting frequency that might facilitate understanding of daily life health behaviors, reduce recall bias if answered promptly, and improve ecological validity and generalizability. Future research using the parent-reported EMA surveys to capture PA and SB of children might use these results to better program their EMA prompts and to also anticipate when parents are more likely to provide more accurate reports of the child's PA and SB. For instance, these studies could incorporate more EMA prompts for the times in which the parent is with the child (eg, evenings hours and weekends) and also incorporate in the study protocol EMA prompts that could be delivered directly to the child, which could be particularly a good strategy for older children.

Sex differences in parent EMA-reported PA and SB and accelerometer-measured minutes of PA and ST were also observed in this sample. Specifically, our results showed a stronger association for boys relative to girls for the relationship between parent EMA-reported PA and accelerometer-measured MVPA. This finding is consistent with previous studies using both objective and self-reported measurements of children's PA, in which showed that boys generally engage in more minutes of MVPA than girls [36-39]. In addition, our results noting sex differences for the association between parent EMA-report of SB and accelerometer-measured ST and LPA are aligned with numerous other studies [23,33,40,41]. It is noteworthy to mention that we observed an inverse association between boys and girls for the association between parent EMA-reported SB and accelerometer-measured ST and LPA. It is possible that parents in our sample perceived boys more often engaged in SB activities (eg, playing video game), whereas girls more often engaged in more LPA (eg, dancing) activities. These findings might be useful for future studies using EMA as a measurement instrument of PA and SB for two main reasons. First, EMA survey question and answer selection could be appropriately designed to reflect specific SB and LPA that are often preferred or habitually practiced by boys (eg, playing video game) and girls (eg, craft, playing musical instruments) [42-44]. Second, these studies could also incorporate children's specific activities preferences for school time, recess time, and leisure time as a way to better capture specific sedentary and LPA behaviors [45,46].

The season has also been noted to be an important factor in modifying children's PA behaviors [25]. However, our results stand in contrast to previous studies examining season effects on the child's PA and SB [47,48], as the association between the parent EMA-reported child's PA and SB and accelerometer-measured ST, LPA, and MVPA remained similar across seasons in our sample. A potential explanation as to why our results differed from past studies may be because of the fact that season was defined differently across studies. For example, some studies defined it as fall versus spring [47], whereas others defined it as warmer versus colder months [48]. This inconsistency in seasonal characterization makes it challenging to compare findings across studies. Therefore, future studies investigating season effects on PA and SB levels in children might benefit from employing more rigorous characterization of seasonality (eg, weather, ecology, hours of daylight, and geographic region).

### Strengths and Limitations

This study has several strengths, including (1) the use of electronically delivered EMA surveys to measure the child's PA and SB, with a flexible prompting frequency delivered repeatedly at different waking hours within an 8-day observation period, which reduced reliance on participant's memory thus reducing the likelihood of recall bias; (2) the use of objective measurement of PA and SB via accelerometry, which is particularly suitable for measuring PA and ST; (3) data collection spanning different seasons and all months of the year; and (4) the inclusion of racially/ethnically and socioeconomically diverse participants, as well as immigrant populations, which increases generalizability within vulnerable/high-risk populations. However, some limitations are noteworthy. First, the dichotomous nature of the EMA survey questions about PA and SB and the lack of clearer descriptions of what constitutes LPA (eg, standing and walking slowly), MVPA (eg, running and biking fast), and SB (eg, time spent watching television or video game) might have led parents to interpret PA and SB differently, which could have resulted in some measurement error. Future studies using EMA surveys might benefit from providing clearer descriptions of what constitutes LPA, MVPA, and SB in their questions, specifically those associated with childhood obesity and other preventable health conditions (eg, viewing television/movie, playing computer and video games, or other screen-related SBs) [49,50]. This would allow parents to provide better reports of their child's behaviors, thus allowing researchers to use more accurate information regarding the specific types of behaviors, such as SBs among children that could be targeted in future behavior intervention trials. Second, it was unknown if parents were in the presence of their child while answering the EMA surveys. Therefore, it is possible that parents were not always observing their child when prompted to answer the EMA surveys and were instead making an educated guess regarding their child's sedentary and PA behaviors, likely because they are knowledgeable about their child's routine [51]. In addition, given that self-report methods of PA (either reported by older children or parent proxy for younger children) might be prone to social desirability bias, we cannot rule out this possibility in our sample. Third, given that accelerometer wear time decreased considerably later in the day and around the time that the fourth signal-contingent EMA survey was delivered to the parent, it is possible that accelerometer-measured child ST, LPA, and MVPA were underestimated during that time of the day. Fourth, our accelerometer wear time data were summarized to the hour, and although responses were uniformly assessed within the hour, future studies should attempt to collect and summarize accelerometer data to at least the minute to gain more precise estimates of PA and ST. Fifth, our results may not be generalizable to activity, and SBs performed outside the times in which the signal-contingent EMA prompts were answered by the parents.

## Conclusions

Our findings indicated that the parent's report of their child's PA and SB via electronically delivered EMA surveys were strongly associated with accelerometer-measured ST, LPA, and MVPA in children aged 5 to 7 years. Notably, these associations were stronger during weekend days than on weekdays. In addition, the parent's EMA report of PA and accelerometer-measured MVPA were more strongly associated in boys relative to girls. The association between the parent's EMA report of SB and accelerometer-measured ST and LPA also differed between boys and girls. Given these findings, in contexts where the parent is able to spend more proximate time observing the child's engagement in PA and SB, the parent's EMA report of their child's PA and SB might be a useful and cost-effective method for measuring PA and SB, particularly in young children and relative to other retrospective self-report measures. Although the concomitant use of EMA and accelerometry is recommended, the use of mobile EMA surveys could be considered when the use of objective measurement of ST and PA are cost prohibitive.

## Acknowledgments

The research was supported by grant number R01HL126171 from the National Heart, Lung, and Blood Institute (principal investigator: JB). A diversity supplement from the parent R01 grant funded the training of the first author (J de Brito). The authors would like to thank the participants and the dedicated staff who carried out the home visits.

## Authors' Contributions

JNdB led the development of study conception and design; conducted data analysis; interpreted the data; wrote the manuscript; and coordinated revisions to the manuscript. KAL assisted with the development of study concept and design; assisted with interpretation of the data; and critically revised the manuscript. AT assisted with the development of study concept and design; conducted data analysis; assisted with interpretation of the data; and assisted with writing and thorough review of the manuscript. JMB is the principal investigator of the Family Matters study; acted as a guarantor of the integrity of the entire Family Matters study; led the development of the Family Matters study concept and design; assisted with data acquisition; and critically revised the present manuscript. All authors approved the final version of the manuscript for submission.

## Conflicts of Interest

None declared.

Multimedia Appendix 1

Association between ecological momentary assessment of the parent-reported child's physical activity and sedentary behavior on accelerometer-measured minutes of sedentary time, light physical activity, and moderate-to-vigorous physical activity per hour in the ecological momentary assessment survey pre-period by day of the week (N=140 for matched ecological momentary assessment accelerometry respondents; N=944 for observation days; and N=3127 for ecological momentary assessment surveys).
[DOCX File , 18 KB - mhealth_v8i5e15458_app1.docx ]

Multimedia Appendix 2

Association between ecological momentary assessment of the parent-reported child's physical activity and sedentary behavior on accelerometer-measured minutes of sedentary time and light and moderate-to-vigorous physical activities per hour in the ecological momentary assessment survey pre-period by sex (N=140 for matched ecological momentary assessment accelerometry respondents; N=944 for observation days; N=3127 for ecological momentary assessment surveys).
[DOCX File , 18 KB - mhealth_v8i5e15458_app2.docx ]

## References

1. Tremblay MS, LeBlanc AG, Kho ME, Saunders TJ, Larouche R, Colley RC, et al. Systematic review of sedentary behaviour and health indicators in school-aged children and youth. Int J Behav Nutr Phys Act 2011 Sep 21;8:98 [FREE Full text] [doi: 10.1186/1479-5868-8-98] [Medline: 21936895]
2. Nelson T, MacLehose R, Davey C, Rode P, Nanney M. Increasing inequality in physical activity among Minnesota secondary schools, 2001-2010. J Phys Act Health 2018 May 1;15(5):325-330. [doi: 10.1123/jpah.2016-0444] [Medline: 29419346]
3. Gordon-Larsen P, Nelson MC, Page P, Popkin BM. Inequality in the built environment underlies key health disparities in physical activity and obesity. Pediatrics 2006 Feb;117(2):417-424. [doi: 10.1542/peds.2005-0058] [Medline: 16452361]
4. US Department of Health and Human Services. Health - US government. Washington, DC: US Department of Health and Human Services; 2018. Physical Activity Guidelines for Americans. Second Edition URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2020-03-17]
5. Matthews CE, Kozey Keadle S, Moore SC, Schoeller DS, Carroll RJ, Troiano RP, et al. Measurement of active and sedentary behavior in context of large epidemiologic studies. Med Sci Sports Exerc 2018 Feb;50(2):266-276 [FREE Full text] [doi: 10.1249/MSS.0000000000001428] [Medline: 28930863]

6. Byrom B, Stratton G, Mc Carthy M, Muehlhausen W. Objective measurement of sedentary behaviour using accelerometers. Int J Obes (Lond) 2016 Nov;40(11):1809-1812 [FREE Full text] [doi: 10.1038/ijo.2016.136] [Medline: 27478922]
7. Sirard JR, Pate RR. Physical activity assessment in children and adolescents. Sports Med 2001;31(6):439-454. [doi: 10.2165/00007256-200131060-00004] [Medline: 11394563]
8. Pate RR, Almeida MJ, McIver KL, Pfeiffer KA, Dowda M. Validation and calibration of an accelerometer in preschool children. Obesity (Silver Spring) 2006 Nov;14(11):2000-2006 [FREE Full text] [doi: 10.1038/oby.2006.234] [Medline: 17135617]
9. Bailey RC, Olson J, Pepper SL, Porszasz J, Barstow TJ, Cooper DM. The level and tempo of children's physical activities: an observational study. Med Sci Sports Exerc 1995 Jul;27(7):1033-1041. [doi: 10.1249/00005768-199507000-00012] [Medline: 7564970]
10. Welk GJ, Corbin CB, Dale D. Measurement issues in the assessment of physical activity in children. Res Q Exerc Sport 2000 Jun;71(2 Suppl):S59-S73. [Medline: 10925827]
11. Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. J Nutr 1988 Nov;118(11):1278-1289. [doi: 10.1093/jn/118.11.1278] [Medline: 3142975]
12. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S68-S76 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399e5b] [Medline: 22157777]
13. Ojiambo RM. Welcome to Enlighten: Theses. Glasgow, Scotland: University of Glasgow; 2012. Assessment of Physical Activity in Children and Adolescents URL: http://theses.gla.ac.uk/3666/1/2012OjiamboPhD.pdf [accessed 2019-05-01]
14. Hidding LM, Chinapaw MJ, van Poppel MN, Mokkink LB, Altenburg TM. An updated systematic review of childhood physical activity questionnaires. Sports Med 2018 Dec;48(12):2797-2842 [FREE Full text] [doi: 10.1007/s40279-018-0987-0] [Medline: 30298479]
15. Klesges LM, Baranowski T, Beech B, Cullen K, Murray DM, Rochon J, et al. Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8- to 10-year-old African-American girls: results from the Girls Health Enrichment Multisite Studies (GEMS). Prev Med 2004 May;38 Suppl:S78-S87. [doi: 10.1016/j.ypmed.2003.07.003] [Medline: 15072862]
16. LeBlanc A, Janssen I. Difference between self-reported and accelerometer measured moderate-to-vigorous physical activity in youth. Pediatr Exerc Sci 2010 Nov;22(4):523-534. [doi: 10.1123/pes.22.4.523] [Medline: 21242602]
17. Kavanaugh K, Moore JB, Hibbett LJ, Kaczynski AT. Correlates of subjectively and objectively measured physical activity in young adolescents. J Sport Health Sci 2015;4(3):222-227. [doi: 10.1016/j.jshs.2014.03.015]
18. Dishman RK, Washburn RA, Schoeller DA. Measurement of physical activity. Quest 2001 Aug;53(3):295-309. [doi: 10.1080/00336297.2001.10491746]
19. Trull TJ, Ebner-Priemer UW. Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section. Psychol Assess 2009 Dec;21(4):457-462 [FREE Full text] [doi: 10.1037/a0017653] [Medline: 19947780]
20. Brockman R, Fox KR, Jago R. What is the meaning and nature of active play for today's children in the UK? Int J Behav Nutr Phys Act 2011 Mar 7;8:15 [FREE Full text] [doi: 10.1186/1479-5868-8-15] [Medline: 21385336]
21. Dunton GF, Liao Y, Intille SS, Spruijt-Metz D, Pentz M. Investigating children's physical activity and sedentary behavior using ecological momentary assessment with mobile phones. Obesity (Silver Spring) 2011 Jun;19(6):1205-1212 [FREE Full text] [doi: 10.1038/oby.2010.302] [Medline: 21164502]
22. Rofey DL, Hull EE, Phillips J, Vogt K, Silk JS, Dahl RE. Utilizing Ecological Momentary Assessment in pediatric obesity to quantify behavior, emotion, and sleep. Obesity (Silver Spring) 2010 Jun;18(6):1270-1272 [FREE Full text] [doi: 10.1038/oby.2009.483] [Medline: 20019675]
23. Zink J, Belcher BR, Dzubur E, Ke W, O'Connor S, Huh J, et al. Association between self-reported and objective activity levels by demographic factors: Ecological momentary assessment study in children. JMIR Mhealth Uhealth 2018 Jun 28;6(6):e150 [FREE Full text] [doi: 10.2196/mhealth.9592] [Medline: 29954723]
24. Hart CN, Raynor HA, Osterholt KM, Jelalian E, Wing RR. Eating and activity habits of overweight children on weekdays and weekends. Int J Pediatr Obes 2011 Oct;6(5-6):467-472 [FREE Full text] [doi: 10.3109/17477166.2011.590204] [Medline: 21774578]
25. Ridley K, Olds T, Hands B, Larkin D, Parker H. Intra-individual variation in children's physical activity patterns: Implications for measurement. J Sci Med Sport 2009 Sep;12(5):568-572. [doi: 10.1016/j.jsams.2008.09.009] [Medline: 19038580]
26. Collings PJ, Wijndaele K, Corder K, Westgate K, Ridgway CL, Dunn V, et al. Levels and patterns of objectively-measured physical activity volume and intensity distribution in UK adolescents: the ROOTS study. Int J Behav Nutr Phys Act 2014 Feb 24;11:23 [FREE Full text] [doi: 10.1186/1479-5868-11-23] [Medline: 24564949]
27. Berge JM, Trofholz A, Tate AD, Beebe M, Fertig A, Miner MH, et al. Examining unanswered questions about the home environment and childhood obesity disparities using an incremental, mixed-methods, longitudinal study design: The Family Matters study. Contemp Clin Trials 2017 Nov;62:61-76 [FREE Full text] [doi: 10.1016/j.cct.2017.08.002] [Medline: 28800894]

28. 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]
29. Evenson KR, Catellier DJ, Gill K, Ondrak KS, McMurray RG. Calibration of two objective measures of physical activity for children. J Sports Sci 2008 Dec;26(14):1557-1565. [doi: 10.1080/02640410802334196] [Medline: 18949660]
30. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
31. Freedman DA. On the so-called 'Huber Sandwich Estimator' and 'Robust Standard Errors'. Am Stat 2006;60(4):299-302. [doi: 10.1198/000313006X152207]
32. Pan W, Connett JE. Selecting the working correlation structure in generalized estimating equations with application to the lung health study. Stat Sin 2002 Apr;12(2):475-490 [FREE Full text]
33. Aznar S, Naylor PJ, Silva P, Pérez M, Angulo T, Laguna M, et al. Patterns of physical activity in Spanish children: a descriptive pilot study. Child Care Health Dev 2011 May;37(3):322-328. [doi: 10.1111/j.1365-2214.2010.01175.x] [Medline: 21083695]
34. Verbestel V, de Henauw S, Bammann K, Barba G, Hadjigeorgiou C, Eiben G, IDEFICS Consortium. Are context-specific measures of parental-reported physical activity and sedentary behaviour associated with accelerometer data in 2-9-year-old European children? Public Health Nutr 2015 Apr;18(5):860-868. [doi: 10.1017/S136898001400086X] [Medline: 24887315]
35. Spruijt-Metz D, Wen CK, Bell BM, Intille S, Huang JS, Baranowski T. Advances and controversies in diet and physical activity measurement in youth. Am J Prev Med 2018 Oct;55(4):e81-e91 [FREE Full text] [doi: 10.1016/j.amepre.2018.06.012] [Medline: 30135037]
36. Janz KF, Burns TL, Levy SM, Iowa Bone Development Study. Tracking of activity and sedentary behaviors in childhood: the Iowa Bone Development Study. Am J Prev Med 2005 Oct;29(3):171-178. [doi: 10.1016/j.amepre.2005.06.001] [Medline: 16168865]
37. Ishii K, Shibata A, Adachi M, Nonoue K, Oka K. Gender and grade differences in objectively measured physical activity and sedentary behavior patterns among Japanese children and adolescents: a cross-sectional study. BMC Public Health 2015 Dec 18;15:1254 [FREE Full text] [doi: 10.1186/s12889-015-2607-3] [Medline: 26679503]
38. Nawrocka A, Mynarski W, Grabara M, Powerska-Didkowska A, Borek Z. Weekday and weekend moderate to vigorous physical activity of young musicians in the context of public health recommendations. Ann Agric Environ Med 2013;20(3):566-570 [FREE Full text] [Medline: 24069866]
39. Trost SG, McCoy TA, Veur SS, Mallya G, Duffy ML, Foster GD. Physical activity patterns of inner-city elementary schoolchildren. Med Sci Sports Exerc 2013 Mar;45(3):470-474. [doi: 10.1249/MSS.0b013e318275e40b] [Medline: 23059861]
40. Cherney ID, London K. Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. Sex Roles 2006;54(9-10):717-726. [doi: 10.1007/s11199-006-9037-8]
41. Verloigne M, van Lippevelde W, Maes L, Yıldırım M, Chinapaw M, Manios Y, et al. Levels of physical activity and sedentary time among 10- to 12-year-old boys and girls across 5 European countries using accelerometers: an observational study within the ENERGY-project. Int J Behav Nutr Phys Act 2012 Mar 31;9:34 [FREE Full text] [doi: 10.1186/1479-5868-9-34] [Medline: 22462550]
42. Gutierrez D, García-López LM. Gender differences in game behaviour in invasion games. Phys Educ Sport Pedagog 2012;17(3):289-301. [doi: 10.1080/17408989.2012.690379]
43. Temple VA, Crane JR, Brown A, Williams B, Bell RI. Recreational activities and motor skills of children in kindergarten. Phys Educ Sport Pedagog 2016;21(3):268-280. [doi: 10.1080/17408989.2014.924494]
44. van Rheenen D. A century of historical change in the game preferences of American children. J Am Folk 2012;125(498):411. [doi: 10.5406/jamerfolk.125.498.0411]
45. Woods AM, Graber KC, Daum DN, Gentry C. Young school children's recess physical activity: Movement patterns and preferences. J Teach Phys Educ 2015;34(3):496-516. [doi: 10.1123/jtpe.2014-0048]
46. Kristjansdottir G, Vilhjalmsson R. Sociodemographic differences in patterns of sedentary and physically active behavior in older children and adolescents. Acta Paediatr 2001 Apr;90(4):429-435. [doi: 10.1111/j.1651-2227.2001.tb00445.x] [Medline: 11332936]
47. Saint-Maurice P, Bai Y, Vazou S, Welk G. Youth physical activity patterns during school and out-of-school time. Children (Basel) 2018 Aug 30;5(9):pii: E118 [FREE Full text] [doi: 10.3390/children5090118] [Medline: 30200255]
48. Silva P, Seabra A, Saint-Maurice P, Soares-Miranda L, Mota J. Physical activity intensities in youth: the effect of month of assessment. Ann Hum Biol 2013;40(5):459-462. [doi: 10.3109/03014460.2013.788211] [Medline: 23829872]
49. Dietz WH. The obesity epidemic in young children. Reduce television viewing and promote playing. Br Med J 2001 Feb 10;322(7282):313-314 [FREE Full text] [doi: 10.1136/bmj.322.7282.313] [Medline: 11159642]
50. Epstein LH, Paluch RA, Gordy CC, Dorn J. Decreasing sedentary behaviors in treating pediatric obesity. Arch Pediatr Adolesc Med 2000 Mar;154(3):220-226. [doi: 10.1001/archpedi.154.3.220] [Medline: 10710017]
51. Board on Children Youth and Families, Division of Behavioral and Social Sciences and Education, National Academies of Sciences Engineering and Medicine, Committee on Supporting the Parents of Young Children, Gadsden V, Ford M, et al. Parenting Matters: Supporting Parents of Children Ages 0-8. Washington, DC: National Academies Press; 2016.

## Abbreviations

**EMA:** ecological momentary assessment
**LPA:** light-intensity physical activity
**MVPA:** moderate-to-vigorous–intensity physical activity
**PA:** physical activity
**SB:** sedentary behavior
**ST:** sedentary time

*Edited by G Eysenbach; submitted 11.07.19; peer-reviewed by T Mason, B Belcher, J Zink; comments to author 31.08.19; revised version received 16.10.19; accepted 26.02.20; published 19.05.20.*

*Please cite as:*
*de Brito JN, Loth KA, Tate A, Berge JM*
*Associations Between Parent Self-Reported and Accelerometer-Measured Physical Activity and Sedentary Time in Children: Ecological Momentary Assessment Study*
*JMIR Mhealth Uhealth 2020;8(5):e15458*
*URL: http://mhealth.jmir.org/2020/5/e15458/*
*doi:10.2196/15458*
*PMID:32348283*

Original Paper

# Mood and Stress Evaluation of Adult Patients With Moyamoya Disease in Korea: Ecological Momentary Assessment Method Using a Mobile Phone App

Yong Sook Yang[1], MSN; Gi Wook Ryu[1], MPH; Chang Gi Park[2], PhD; Insun Yeom[3], PhD; Kyu Won Shim[3*], MD, PhD; Mona Choi[1*], PhD

[1]Mo-Im Kim Nursing Research Institute, Yonsei University College of Nursing, Seoul, Republic of Korea

[2]College of Nursing, University of Illinois at Chicago, Chicago, IL, United States

[3]Department of Pediatric Neurosurgery, Yonsei University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

**Corresponding Author:**
Mona Choi, PhD
Mo-Im Kim Nursing Research Institute
Yonsei University College of Nursing
50 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 222283341
Fax: 82 222278303
Email: monachoi@yuhs.ac

## Abstract

**Background:** Moyamoya disease (MMD) is a known progressive obstructive cerebrovascular disorder. Monitoring and managing mood and stress are critical for patients with MMD, as they affect clinical outcomes. The ecological momentary assessment (EMA) method is a longitudinal study design by which multiple variable assessments can be performed over time to detect momentary fluctuations and changes in psychological dimensions such as mood and stress over time.

**Objective:** This study aimed to identify predicting factors associated with momentary mood and stress at both the within-person and between-person levels and to examine individual fluctuation of mood over time in the short term using an EMA method combined with a mobile phone app.

**Methods:** Participants aged older than 18 years were recruited from a tertiary hospital in Seoul, Korea, between July 2018 and January 2019. The PsyMate scale for negative affect (NA) and positive affect (PA) and the Trier Inventory for Chronic Stress Scale were uploaded on patient mobile phones. Using a mobile app, data were collected four times a day for 7 days. Pearson correlations and mixed modeling were used to predict relationships between repeatedly measured variables at both the between-person and within-person levels.

**Results:** The mean age of the 93 participants was 40.59 (SD 10.06) years, 66 (71%) were female, and 71 (76%) were married. Participants provided 1929 responses out of a possible 2604 responses (1929/2604, 74.08%). The mean momentary NA and PA values were 2.15 (SD 1.12) and 4.70 (SD 1.31) out of 7, respectively. The momentary stress value was 2.03 (SD 0.98) out of 5. Momentary NA, PA, and stress were correlated ($P$<.001) and varied over time in relation to momentary variables. Common momentary variables associated with momentary mood and stress at both the within-person (level 1) and between-person (level 2) levels were identified. Momentary NA increased when being alone and being at the hospital at both levels, whereas momentary PA increased when eating or drinking, resting, being at a café, restaurant or a public place but decreased when being alone at both levels. Momentary stress increased when being at the office, at a public place, or as the time of the day went by but decreased when resting or during the weekend. Different factors affecting mood and stress at different levels were identified. Fluctuations in individual momentary mood over time at the within-person level were captured.

**Conclusions:** The EMA method using a mobile phone app demonstrated its ability to capture changes in mood and stress in various environmental contexts in patients with MMD. The results could provide baseline information for developing interventions to manage negative mood and stress of patients with MMD based on the identified predictors affecting mood and stress at two different levels.

*(JMIR Mhealth Uhealth 2020;8(5):e17034)* doi:10.2196/17034

**KEYWORDS**

affect; ecological momentary assessment; mood; Moyamoya disease; psychological stress

# Introduction

## Background

Moyamoya disease (MMD) is a rare idiopathic vascular disorder that is characterized by progressive bilateral stenosis or occlusion of the distal branches of the carotid arteries with an abnormal vascular network [1,2]. It is associated with the development of bifurcation and compensatory arterial collateral networks at the base of the brain [1,2], hence the name *moyamoya* (describing a puff of smoke in Japanese) [1]. MMD is common in Asian countries, such as Korea and Japan, with sparse observation in Europe and the Americas [3,4].

The prevalence of MMD in Korea has gradually increased and reached 16.1 per 100,000 persons in 2011 [4]. This increase can be partly explained by the increasing availability of diagnostic tests, such as magnetic resonance (MR) imaging or MR angiography and prolonged survival owing to improved management [4,5]. In Japan, the prevalence rate was 6.03 per 100,000 persons in 2003 [6], and in Nanjing, China, the prevalence rate was 3.92 per 100,000 persons in 2010 [7]. The incidence rate of MMD in the United States was lower than that in East Asian countries at 0.087 per 100,000 persons [8], whereas moyamoya syndrome (MMS), termed unilateral moyamoya angiopathy [9], is considered to have higher prevalence in Western countries than in East Asian countries [10]. The prevalence of MMS in Western countries was found to be close to that in Japan at 0.34 per 100,000 persons in a nationwide survey [11]. The relatively high prevalence of MMS in Western countries may be related to the fact that sickle cell disease is a frequent cause of MMS in individuals of African origin living in the United States and Europe [12,13].

The clinical features of MMD or MMS in adult patients generally involve cerebral hemorrhages and infarction, whereas children develop ischemic attacks [4]. Relating to the disease's chronic and uncertain nature, it has been observed that improper management of mood and stress can lead to reduced cerebrovascular blood flow, which is closely related to the prognosis of the disease [14]. According to recent studies, adults with MMD are vulnerable to stress and mood changes [14] and may have anxiety, depression, and posttraumatic stress syndrome [15].

Mood is a state of subjective feeling that can be changed by events and is typically described as having either positive or negative valences [16]. Individual differences exist in experiencing the states of positive and negative feelings, and these are assessed as positive and negative effects [17]. Negative mood states such as depression can extend to psychosocial distress that potentially leads to stress and, more generally, to a negative outlook on life [18,19]. Mood and stress vary with time in relation to the surrounding context, showing changes over time and interindividual differences [20].

The ecological momentary assessment (EMA) method, also known as the experience sample method or ambulatory assessment, is a repeated observational study design by which time-varying variables can be assessed in natural and real-life environments [20-22]. For example, participants self-report their mood or anxiety multiple times to observe changes based on the environment and/or time while performing their ordinary daily tasks, without changing their life patterns to attend a survey [23]. Participants can also report various events, types of food, and calories per serving multiple times per day to track down changes in their eating behavior over days or weeks [24]. Such reports can be completed through diaries, personal digital assistants, mobile apps, or wearable sensors [25] in the short or long term, depending on the study goals and design [26]. This method is reportedly accurate and able to detect changes in psychological properties through multiple daily assessments [27,28], helps address questions regarding individual differences [29], and elucidates momentary changes and fluctuations in psychological dimensions such as mood and stress over time and across situations [30,31]. It has been widely used to assess the psychological characteristics of participants with [32,33] and without [26] mental problems. Recently, various tools using mobile phone technology have been developed and used to measure mood and stress in diverse patient populations [26]; thus, this study utilized the EMA approach for moyamoya patients' condition.

## Objectives

This study aimed to identify predicting factors associated with momentary mood and stress at both the within-person and between-person levels and to examine individual fluctuation of mood over time in the short term using an EMA method combined with a mobile phone app.

# Methods

## Participants

Adult patients with MMD, who visited the outpatient clinic of a tertiary hospital or were admitted in the same hospital, were recruited from July 2018 to January 2019. Only participants who used Android operating systems were included, as the developed mobile app was only available for this operating system with the version 4.4 or higher as described in a previous study [32]. To exclude possible cognitive impairment that would preclude the participants from answering the questionnaire, all patients were required to have a score higher than 24 on the Korean version of the Mini-Mental State Examination [33], which is widely used to test the cognitive ability of clinical populations in Korea. This study was approved by the institutional review board of the Yonsei University Health System (approval number: 4-2018-0385), and informed consent was obtained from each participant.

## Measurements

Measurements in this study included baseline variables, such as demographic characteristics, disease-specific information, trait mood (anxiety and depression), and trait stress. Momentary measures were mood and stress. The study variables at baseline and momentary measures are summarized in Table 1.

**Table 1.** Study variables at all time points.

| Variables | Baseline | Ecological momentary assessment | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Day 1 | Day 2 | Day 3 | Day 4 | Day 5 | Day 6 | Day 7 |
| Demographic characteristics | x | N/A[a] | N/A | N/A | N/A | N/A | N/A | N/A |
| Disease-specific information | x | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Korean Hospital Anxiety and Depression Scale | x | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Korean Perceived Stress Scale | x | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| Momentary mood (PsyMate) | N/A | x | x | x | x | x | x | x |
| Momentary stress (Trier Inventory for Chronic Stress) | N/A | x | x | x | x | x | x | x |

[a]x: variables measured at the day point.

[b]N/A: not applicable.

## Baseline Measures

Patients provided demographic information such as age, sex, income, level of education, symptoms experienced, and disease duration since the diagnosis. The perceived severity of the disease was also self-reported at baseline using a 5-point scale.

We used the Korean version of the Hospital Anxiety and Depression Scale (K-HADS) [34] to measure trait mood after obtaining the permission of the scale provider. The HADS is known as a reliable and valid scale and used worldwide for measuring mental health in clinical settings [35]. The Cronbach alphas of the K-HADS anxiety and depression subscales have been reported to be .89 and .86, respectively [35]. The scale consists of 14 items, 7 items for assessing anxiety and 7 items for depression, measured on a 4-point scale, from 0 to 3. A higher score denotes a higher level of anxiety or depression. The Cronbach alphas for the anxiety and depression subscales in this study were .75 and .81, respectively.

To measure trait stress, we used the Korean version of the Perceived Stress Scale, which consists of 10 items [36]. This scale was freely downloaded from the official homepage of the Laboratory for the Study of Stress, Immunity, and Disease of the Carnegie Mellon University. It has negative and positive subdomains and five negative items for stress and five positive items for coping ability rated on a 5-point Likert scale (0–4). The Cronbach alphas of the two subdomains have been reported to be .87 and .71, respectively [36]. The Cronbach alpha in this study was .89 for both subdomains.

## Momentary Measures Using a Mobile App

Momentary mood was measured using the Korean version of the PsyMate, translated from the English version with reference to the original Dutch version. We obtained permission to use PsyMate from the developers [37]. It consists of nine items assessing negative affect (NA) and four items assessing positive affect (PA). The Cronbach alphas of the subscales of NA and PA were .91 and .92, respectively, in the previous study [37]. The Cronbach alphas in this study were .94 and .92 for NA and PA, respectively.

Stress was assessed by the Korean version of the Trier Inventory for Chronic Stress, adapted from the German version [38,39]. It consists of eight items measuring work overload, social overload, pressure to perform, work discontent, excessive demands from work, lack of social recognition, social tensions, and social isolation. The Cronbach alpha in this study was .79.

The questionnaires were uploaded on a mobile app for the Android operating system developed in the previous study [32]. Momentary mood and stress were measured in an environmental context considering what the participants were doing, where they were, and with whom they were at the moment of answering.

## Procedure

After obtaining participant informed consent, we held an individual and face-to-face 30-min intake session with each patient. Patients filled the baseline measures and were allowed time to download the app and practice answering for the EMA study. Researchers helped the participants answer the baseline survey and install the app. Patients were provided a reward in coupons when they completed the baseline survey and enrollment. They were also informed that they would receive additional coupons on completion of the EMA study. Researchers preset the survey period for each patient in advance to push notifications and instructed participants to carry their mobile phone during the scheduled survey period and to answer the survey question when they received the notification requesting them to do so.

Measures of mood and stress were set on the mobile app, and notifications were set to appear four times a day for 7 consecutive days (4 times ×7 days=28 times/person) in semirandom, 90-min blocks. Notifications were sent in the morning between 8 AM and 9 PM, early afternoon between noon and 1 PM, evening between 5 PM and 6 PM, and at night between 9 PM and 10 PM. Participants were instructed that they would receive a reminder notification when they did not input the response within 45 min after they received the first notification for each scheduled measurement. Researchers monitored participant compliance to the protocol and managed

participation by phoning patients who did not respond on the first day and attempted to solve any participation difficulties and problems in the EMA study.

### Statistical Analysis

Data analysis was performed with STATA 14.0 (StataCorp) using 1929 responses from 93 participants who provided more than three responses in the total course of the study to capture changes over time [20]. We analyzed participant characteristics by descriptive analysis. Independent *t* tests and analysis of variance were used to compare the mean differences between the variable groups. We estimated the Pearson coefficient to examine the correlation between the momentary variables. Mixed modeling analysis, handling the clustered and correlated data [20], was used to describe and predict factors associated with momentary NA, PA, and stress at both the within-person (level 1) and between-person (level 2) levels. The threshold of statistical significance was set at $P<.05$.

## Results

### Participants

A total of 93 participants with MMD were recruited. Of 93 participants, 71 (76%) were recruited from the outpatient department and 22 participants (24%) were recruited from the admission wards of a university hospital. The mean age of the participants was 40.59 years (SD 10.06), 71% (66/93) participants were female, and 76% (71/93) participants were married. The mean number of years since diagnosis was 3.68 (SD 4.05), and the perceived severity level was 3.56 (SD 1.00) out of 5. The mean HADS anxiety and depression scores were 7.17 (SD 3.38) and 7.14 (SD 3.51) out of 21, respectively, and the mean perceived stress level was 1.64 (SD .98) out of 4. The participant baseline characteristics are summarized in Table 2.

**Table 2.** Participant characteristics at baseline (n=93).

| Characteristics | n (%) | Mean (SD) | Possible range |
|---|---|---|---|
| **Age (years)** | | | |
| 20–29 | 11 (12) | 40.59 (10.06) | N/A[a] |
| 30–39 | 37 (40) | N/A | N/A |
| 40–49 | 26 (28) | N/A | N/A |
| 50–59 | 15 (16) | N/A | N/A |
| ≥60 | 4 (4) | N/A | N/A |
| **Sex** | | | |
| Female | 66 (71) | N/A | N/A |
| Male | 27 (29) | N/A | N/A |
| **Marital status** | | | |
| Married | 71 (76) | N/A | N/A |
| Not married | 22 (24) | N/A | N/A |
| **Education** | | | |
| ≤High school | 41 (44) | N/A | N/A |
| ≥College | 52 (56) | N/A | N/A |
| **Monthly household income[b] (US $)** | | | |
| <2000 | 23 (25) | N/A | N/A |
| 2000–3000 | 16 (18) | N/A | N/A |
| 3000–4000 | 22 (24) | N/A | N/A |
| >4000 | 31 (33) | N/A | N/A |
| Years since the diagnosis | N/A | 3.68 (4.05) | N/A |
| Perceived severity | N/A | 3.56 (1.00) | 1–5 |
| HADS[c] anxiety | N/A | 7.17 (3.38) | 0–21 |
| HADS depression | N/A | 7.14 (3.51) | 0–21 |
| Perceived stress | N/A | 1.64 (.98) | 0–4 |

[a]N/A: not applicable.

[b]n=92.

[c]HADS: Hospital Anxiety and Depression Scale.

## Momentary Responses

Participants provided 1929 responses out of a possible 2604 (1929/2604, 74.1%). Of 1883 responses, 799 (42.4%) were answered while participants were resting, 344 responses (18.3%) while working, 293 responses (15.6%) doing household work, and 188 responses (10.0%) while eating or drinking at the moment of answering. Of the 1929 responses, 1147 (59.5%) were obtained when participants were at home, 325 responses (16.9%) were obtained at the office, and 87 responses (4.5%) were obtained at a café or restaurant. Of all 1929 responses, 444 (23.1%) were obtained while the participants were alone, and 1348 (69.9%) were obtained on weekdays. Of 1929 responses, 489 (25.3%), 497 (25.8%), 509 (26.4%), and 434 (22.5%) were provided in the morning, afternoon, evening, and at night, respectively. The mean momentary NA and PA were 2.15 (SD 1.12) and 4.70 (SD 1.31) out of 7, respectively. The mean momentary stress level was 2.03 (SD .68) out of 5. Measures of momentary variables are summarized in Table 3.

**Table 3.** Measures of momentary variables (n=1929).

| Momentary variables | n (%) | Mean (SD) | Possible range |
|---|---|---|---|
| **What[a] (things doing)** | | | |
| Household work | 293 (15.6) | N/A[b] | N/A |
| Working | 344 (18.3) | N/A | N/A |
| Eating/drinking | 188 (10.0) | N/A | N/A |
| Resting | 799 (42.4) | N/A | N/A |
| Other | 259 (13.7) | N/A | N/A |
| **Where[c] (place being)** | | | |
| Home | 1147 (59.5) | N/A | N/A |
| Office | 325 (16.9) | N/A | N/A |
| Café or restaurant | 87 (4.5) | N/A | N/A |
| Hospital | 66 (3.4) | N/A | N/A |
| Public place | 138 (7.2) | N/A | N/A |
| Other | 164 (8.5) | N/A | N/A |
| **Being alone** | | | |
| Yes | 444 (23.1) | N/A | N/A |
| No | 1485 (76.9) | N/A | N/A |
| **Weekend** | | | |
| Yes | 581 (30.1) | N/A | N/A |
| No | 1348 (69.9) | N/A | N/A |
| **Time of day** | | | |
| Morning (8 AM to 9 PM) | 489 (25.3) | N/A | N/A |
| Afternoon (12 noon to 1 PM) | 497 (25.8) | N/A | N/A |
| Evening (5 PM to 6 PM) | 509 (26.4) | N/A | N/A |
| Night (9 PM to 10 PM) | 434 (22.5) | N/A | N/A |
| Momentary negative affect | N/A | 2.15 (1.12) | 1–7 |
| Momentary positive affect | N/A | 4.70 (1.31) | 1–7 |
| Momentary stress | N/A | 2.03 (0.68) | 1–5 |

[a]n=1883.

[b]N/A: not applicable.

[c]n=1927.

## Correlations Between Momentary Negative Affect, Positive Affect, and Stress

Momentary NA, PA, and stress were significantly correlated ($P$<.001). Momentary NA was negatively correlated with momentary PA ($r$=−0.607; $P$<.001) and positively with momentary stress ($r$=0.538; $P$<.001). Momentary PA was negatively correlated with momentary stress ($r$=−0.272; $P$<.001). The correlation coefficients between variables are presented in Table 4.

**Table 4.** Correlation coefficients between momentary variables (n=93).

| Momentary variables | Momentary NA[a] (r value) | Momentary PA[b] (r value) |
|---|---|---|
| Momentary NA | 1 | -0.607[c] |
| Momentary PA | −0.607[c] | 1 |
| Momentary stress | 0.538[c] | −0.272[c] |

[a]NA: negative affect.

[b]PA: positive affect.

[c]*P*<.001.

### Modeling of Between-Person and Within-Person Analysis

Analysis by mixed modeling was performed based on 1929 completed assessments from the participants who provided more than three responses for analyzing changes of momentary mood and stress over time at both level 1 (within-person) and level 2 (between-person).

Momentary predicting factors were examined by three models at levels 1 and 2. Disease-specific variables such as perceived severity and years since the diagnosis, age, sex, and trait mood variables at the baseline were added into models 2 and 3 for adjustment. Momentary variables included *things doing* (what), *place being* (where), *being alone* or not (with whom), answering during the *weekend* or not, and *time of day*. Table 5 shows the variables included in each model at both the within-person and between-person levels.

**Table 5.** Designed levels, models, and variables.

| Level and model | Variables |
|---|---|
| **Within-person** | |
| 1. Momentary variables | What, where, with whom, weekend, and time of day |
| 2. Disease-specific variables | Variables of Model 1 + perceived severity and years since the diagnosis |
| 3. Trait variables | Variables of Model 2 + age, sex, trait anxiety/depression, and stress |
| **Between-person** | |
| 1. Momentary variables | What, where, with whom, weekend, and time of day |
| 2. Disease-specific variables | Variables of Model 1 + perceived severity and years since the diagnosis |
| 3. Trait variables | Variables of Model 2 + age, sex, trait anxiety/depression, and stress |

### Momentary Variables Affecting Momentary Mood and Stress at the Within-Person and Between-Person Levels

Common momentary variables associated with momentary mood and stress at both the within-person (level 1) and between-person (level 2) levels were identified. Momentary NA increased when being alone and being at the hospital at both levels, whereas momentary PA increased when eating or drinking, resting, being at a café or restaurant, or at the public place but decreased when being alone at both levels. Momentary stress increased when at the office, at the public place, or as the time of the day went by but decreased when resting or during the weekend.

Different factors affecting momentary mood and stress at different levels were also identified. Variables of being at a café or restaurant (coefficient=−0.19; *P*=.03) and during the weekend (coefficient=−.08; *P*=.03) were associated with momentary NA at the within-person level only. However, eating or drinking (coefficient=−0.20; *P*=.04) and resting (coefficient=−0.21; *P*=.01) were associated with momentary NA at the between-person level only. There was no difference in factors associated with momentary PA either at the within-person or at the between-person level. Among variables, being at the hospital (coefficient=5.67; *P*<.001) was associated with momentary stress at the between-person level but not at the within-person level. Table 6 shows the parameter estimates from the mixed effect model at both levels.

**Table 6.** Fixed effect model parameter estimates at within-person and between-person levels in model 3.

| Variables[a] | Momentary negative affect, coefficient (SE) | | Momentary positive affect, coefficient (SE) | | Momentary stress, coefficient (SE) | |
|---|---|---|---|---|---|---|
| | Within | Between | Within | Between | Within | Between |
| **What (things doing)** | | | | | | |
| Household work | Reference | Reference | Reference | Reference | Reference | Reference |
| Working | 0.05 (0.09) | 0.08 (0.12) | −0.20 (0.12) | −0.20 (0.15) | 0.26 (0.42) | 0.64 (0.61) |
| Eating or drinking | −0.09 (0.07) | −0.20 (0.10)[b] | 0.24 (0.09)[b] | 0.27 (0.12)[b] | 0.23 (0.33) | −0.56 (0.49) |
| Resting | −0.08 (0.05) | −0.21 (0.07)[c] | 0.15 (0.06)[b] | 0.30 (0.08)[c] | −1.03 (0.23)[c] | −1.56 (0.33)[c] |
| Other | −0.01 (0.06) | −0.02 (0.09) | 0.03 (0.09) | −0.05 (0.11) | −0.26 (0.31) | −1.20 (0.45) |
| **Where (place being)** | | | | | | |
| Home | Reference | Reference | Reference | Reference | Reference | Reference |
| Office | 0.05 (0.08) | −0.19 (0.12) | −0.04 (0.11) | 0.21 (0.14) | 1.82 (0.41)[c] | 1.56 (0.58)[c] |
| Café or restaurant | −0.19 (0.09)[b] | −0.10 (0.12) | 0.37 (0.12)[c] | 0.33 (0.15)[b] | −0.75 (0.42) | −0.17 (0.61) |
| Hospital | 0.32 (0.12)[b] | 0.54 (0.12)[c] | −0.17 (0.16) | −0.07 (0.15) | 0.60 (0.58) | 5.67 (0.62)[c] |
| Public place | −0.06 (0.07) | 0.01 (0.09) | 0.21 (0.09)[b] | 0.28 (0.12)[b] | 0.91 (0.33)[c] | 2.13 (0.47)[c] |
| Other | −0.10 (0.06) | −0.10 (0.09) | 0.22 (0.08) | 0.35 (0.11) | 0.08 (0.30) | 0.83 (0.43) |
| Being alone | 0.10 (0.04)[b] | 0.24 (0.05)[c] | −0.16 (0.05)[c] | −0.49 (0.07)[c] | −0.37 (0.20) | −0.34 (0.27) |
| Weekend | −0.08 (0.03)[b] | −0.03 (0.05) | 0.02 (0.05) | 0.01 (0.06) | −1.20 (0.16)[c] | −0.76 (0.25)[c] |
| **Time of day** | | | | | | |
| 8-9 AM | Reference | Reference | Reference | Reference | Reference | Reference |
| 12 noon-1 PM | −0.03 (0.04) | −0.01 (0.06) | 0.09 (0.06) | 0.08 (0.08) | 0.76 (0.20)[c] | 0.77 (0.31)[c] |
| 5-6 PM | −0.04 (0.04) | −0.01 (0.06) | 0.03 (0.05) | −0.03 (0.08) | 1.07 (0.20)[c] | 1.03 (0.30)[c] |
| 9-10 PM | −0.04 (0.04) | −0.05 (0.06) | 0.17 (0.06)[c] | 0.14 (0.08) | 1.48 (0.21)[c] | 1.40 (0.32)[c] |
| Perceived severity | 0.11 (0.08) | 0.13 (0.02)[c] | −0.03 (0.09) | −0.06 (0.03)[b] | 0.69 (0.41) | 0.73 (0.11)[c] |
| Years since the diagnosis | −0.03 (0.02) | −0.01 (0.01) | −0.03 (0.02) | −0.04 (0.01)[b] | −0.16 (0.10) | −0.10 (0.03)[c] |
| Trait anxiety and depression | 0.06 (0.02)[c] | 0.07 (0.01)[c] | −0.06 (0.02)[c] | −0.08 (0.01)[c] | 0.21 (0.08)[b] | 0.19 (0.02)[c] |
| Trait stress | 0.04 (0.02)[b] | 0.03 (0.01)[c] | 0.01 (0.02) | 0.02 (0.01)[b] | 0.22 (0.12) | 0.23 (0.03)[c] |

[a]Age and sex adjusted.

[b]$P<.05$.

[c]$P<.01$.

## Individual Fluctuation of Negative Affect and Positive Affect Over Time

Both the momentary NA and PA of participants fluctuated over time. We arbitrarily selected 10 participants from those who completed all 28 assessments to show the individual fluctuation of affect over time. Specific graphs of NA and PA fluctuation were constructed for each selected participant in accordance with momentary and trait variables (Figures 1 and 2 as examples).

**Figure 1.** Individual fluctuation of negative affect over time for the selected participants. NA: negative affect; A to I: selected participants.

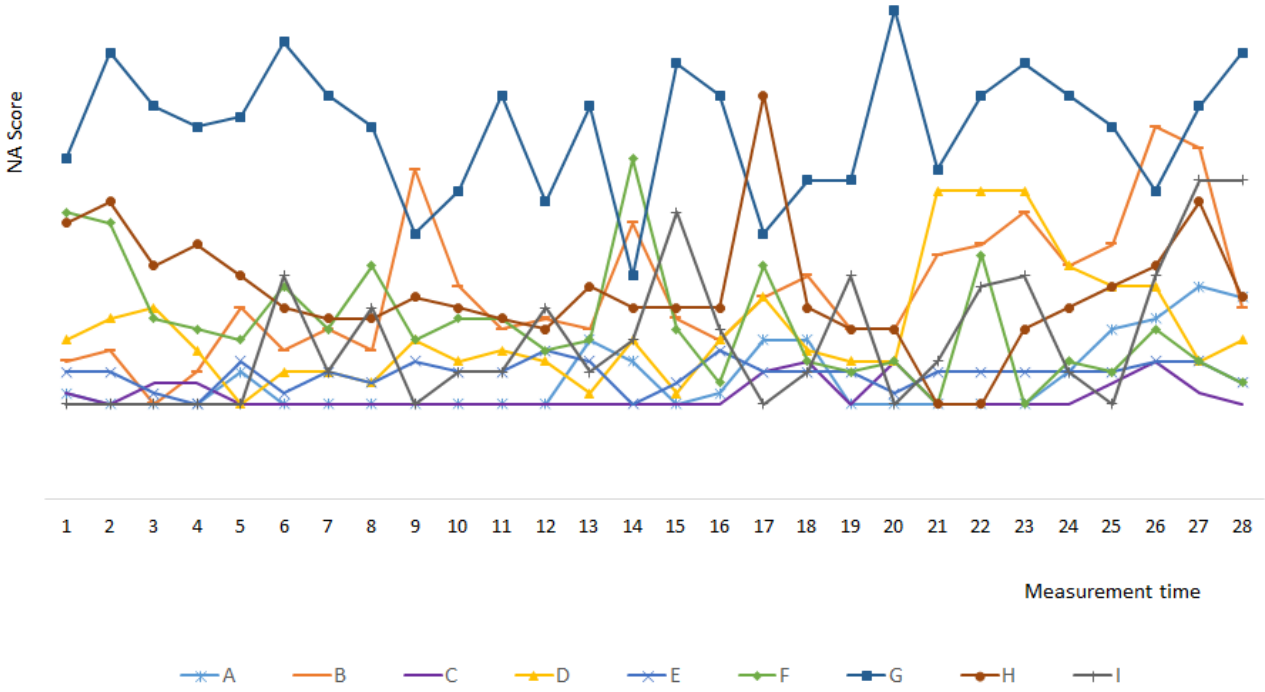
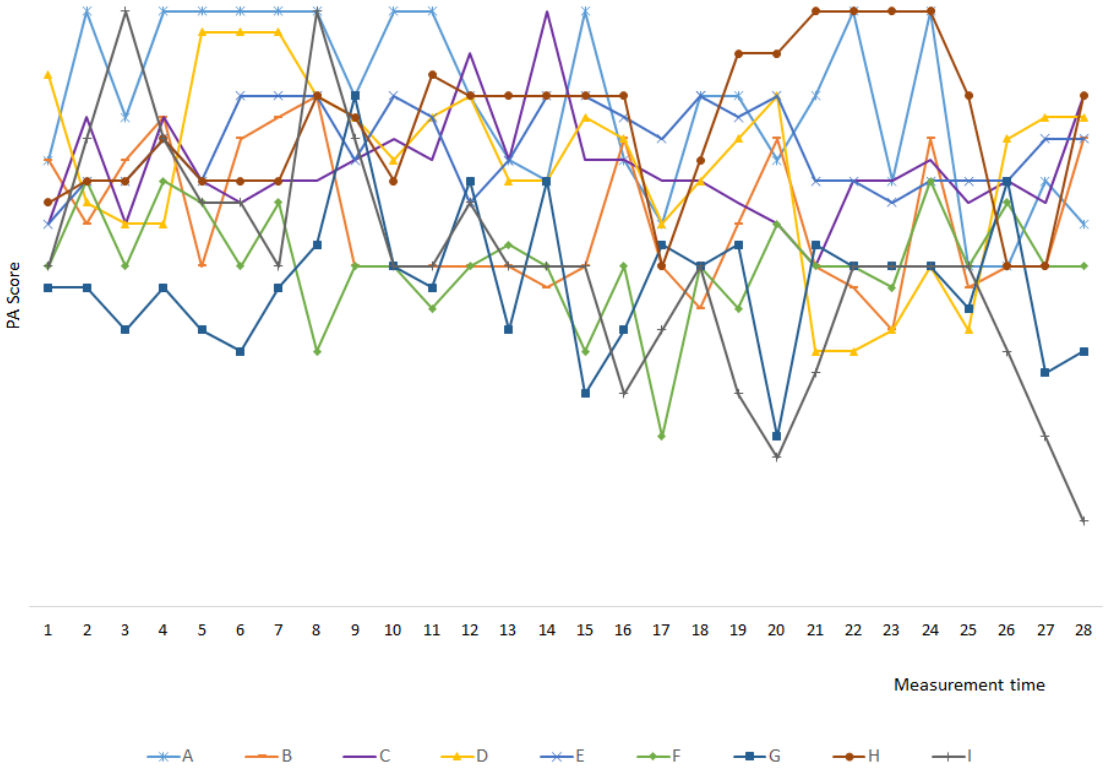



**Figure 2.** Individual fluctuation of positive affect over time for the selected participants. PA: positive affect; A to I: selected participants.




## Discussion

### Principal Findings

This study used the EMA method to assess and predict momentary factors associated with momentary mood and stress in a real-life context in adult patients with MMD. The results showed that context variables of the participants' natural environment affected momentary mood and stress at both the within-person and between-person levels after adjusting for participant demographics and disease-specific characteristics and trait anxiety and depression.

We also identified common variables affecting momentary mood and stress at both the within-person and between-person levels and distinguished different variables having different effects at the within-person and between-person levels. Participants commonly showed higher NA when being alone

or at the hospital at both levels. Meanwhile, some factors such as activities of eating or drinking or resting affected momentary NA at the between-person level only. Participants commonly expressed decreased stress when resting or during the weekend and increased stress at the office and as the time of the day went by at both levels. Meanwhile, being at the hospital was associated with higher stress only at the between-person level. These results showed that different factors at different levels are associated with NA and stress, although there are numerous common factors. These factor differences between the two levels should be considered when designing individualized interventions to manage the mood and stress of patients with MMD.

This study also established that an EMA method using a mobile app could capture individual fluctuations of mood and stress over time while participants perform their usual daily tasks. This result is aligned with the result from a previous study that applied an EMA app using the same scale of the PsyMate to assess the mood of Dutch patients in an ambulatory mental health setting [38], presenting the ability to detect changes in mood over time.

In an additional analysis, we found that patients who had been diagnosed less than a year ago appeared to be more depressed than those who were diagnosed more than a year ago. These results indicate that an emotional care plan with close, regular monitoring is needed for patients with MMD, especially for those with higher anxiety and depression, and within a year after diagnosis. The levels of anxiety and depression of the participants in this study at baseline were 7.17 (SD 3.38) and 7.14 (SD 3.51) out of 21, respectively. This implies that adults with MMD may also strive to overcome negative feelings, as do patients with other cerebrovascular diseases who are at continuous risk of cerebrovascular hemorrhage and infarction [40,41].

Perceived stress is a known predictor of depression and depressive symptoms in patients with stroke [42,43]. The results of this study also showed that both trait and momentary stress are significantly associated with momentary mood at both the within-person and between-person levels. Stress should be managed, as it triggers negative mood. Negative mood and stress in patients with cerebrovascular disorders are related to emotional distress, which threatens health behaviors and drives patients to avoid health-promoting activities [44].

Perceived social support plays a critical role in buffering stress and promoting psychological well-being [45-47]. Patients with MMD should be encouraged to engage in social interactions with family or self-support groups, as it was shown that being alone is significantly related to momentary increase in negative mood at both levels.

### Limitations and Future Directions

Our study had limitations. This study included patients who used Android OS, and those who used other systems were excluded. In addition, there might be challenges regarding technical issues and potential malfunctioning of the configuration, although a helpline was provided by our research team. These technical points need to be addressed to improve the EMA survey in the future. In addition, as the patients were recruited from a tertiary-level hospital in Seoul, patients in communities or in smaller facilities may differ from this study population in terms of clinical severity, years since the diagnosis, or trait mood and stress levels.

Studies on the impact of social support or stress-coping strategies on mood and stress in MMD warrants further investigation. EMA methods integrated into momentary interventions for improving mood and stress could be a promising future direction in MMD.

### Conclusions

In this study, we evaluated the EMA method using a mobile phone app and demonstrated that the EMA method was able to capture mood and stress change over time and by assessing momentary contextual variables. With the identified predictors affecting mood and stress at two different levels, the results of this study could provide valuable information for developing individualized patient-centered interventions for managing the mood and stress of patients with MMD who are psychologically vulnerable and whose states cannot be easily assessed in a real-world environment.

### Acknowledgments

The authors thank the participants of this study for sharing a part of their life. This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2017R1D1A1B03030706).

### Conflicts of Interest

None declared.

### References

1. Suzuki J, Takaku A. Cerebrovascular 'moyamoya' disease. Disease showing abnormal net-like vessels in base of brain. Arch Neurol 1969 Mar;20(3):288-299. [doi: 10.1001/archneur.1969.00480090076012] [Medline: 5775283]
2. Natori Y, Ikezaki K, Matsushima T, Fukui M. 'Angiographic moyamoya' its definition, classification, and therapy. Clin Neurol Neurosurg 1997 Oct;99(Suppl 2):S168-S172. [doi: 10.1016/s0303-8467(97)00052-8] [Medline: 9409430]
3. Huang S, Guo Z, Shi M, Yang Y, Rao M. Etiology and pathogenesis of Moyamoya Disease: An update on disease prevalence. Int J Stroke 2017 Apr;12(3):246-253. [doi: 10.1177/1747493017694393] [Medline: 28381201]

4. Ahn IM, Park D, Hann HJ, Kim KH, Kim HJ, Ahn HS. Incidence, prevalence, and survival of moyamoya disease in Korea: a nationwide, population-based study. Stroke 2014 Apr;45(4):1090-1095. [doi: 10.1161/STROKEAHA.113.004273] [Medline: 24595588]
5. Kim JS. Moyamoya disease: epidemiology, clinical features, and diagnosis. J Stroke 2016 Jan;18(1):2-11 [FREE Full text] [doi: 10.5853/jos.2015.01627] [Medline: 26846755]
6. Kuriyama S, Kusaka Y, Fujimura M, Wakai K, Tamakoshi A, Hashimoto S, et al. Prevalence and clinicoepidemiological features of moyamoya disease in Japan: findings from a nationwide epidemiological survey. Stroke 2008 Jan;39(1):42-47. [doi: 10.1161/STROKEAHA.107.490714] [Medline: 18048855]
7. Miao W, Zhao P, Zhang Y, Liu H, Chang Y, Ma J, et al. Epidemiological and clinical features of Moyamoya disease in Nanjing, China. Clin Neurol Neurosurg 2010 Apr;112(3):199-203. [doi: 10.1016/j.clineuro.2009.11.009] [Medline: 20004511]
8. Uchino K, Johnston SC, Becker KJ, Tirschwell DL. Moyamoya disease in Washington State and California. Neurology 2005 Sep 27;65(6):956-958. [doi: 10.1212/01.wnl.0000176066.33797.82] [Medline: 16186547]
9. Phi JH, Wang K, Lee JY, Kim S. Moyamoya syndrome: a window of Moyamoya disease. J Korean Neurosurg Soc 2015 Jun;57(6):408-414 [FREE Full text] [doi: 10.3340/jkns.2015.57.6.408] [Medline: 26180607]
10. Hitomi T, Habu T, Kobayashi H, Okuda H, Harada KH, Osafune K, et al. Downregulation of Securin by the variant RNF213 R4810K (rs112735431, G>A) reduces angiogenic activity of induced pluripotent stem cell-derived vascular endothelial cells from moyamoya patients. Biochem Biophys Res Commun 2013 Aug 16;438(1):13-19. [doi: 10.1016/j.bbrc.2013.07.004] [Medline: 23850618]
11. Hayashi K, Horie N, Suyama K, Nagata I. An epidemiological survey of moyamoya disease, unilateral moyamoya disease and quasi-moyamoya disease in Japan. Clin Neurol Neurosurg 2013 Jul;115(7):930-933. [doi: 10.1016/j.clineuro.2012.09.020] [Medline: 23041378]
12. Guey S, Tournier-Lasserve E, Hervé D, Kossorotoff M. Moyamoya disease and syndromes: from genetics to clinical management. Appl Clin Genet 2015;8:49-68 [FREE Full text] [doi: 10.2147/TACG.S42772] [Medline: 25733922]
13. Fryer RH, Anderson RC, Chiriboga CA, Feldstein NA. Sickle cell anemia with moyamoya disease: outcomes after EDAS procedure. Pediatr Neurol 2003 Aug;29(2):124-130. [doi: 10.1016/s0887-8994(03)00047-x] [Medline: 14580655]
14. Hayashi S, Terada S, Sato S, Oshima E, Miki T, Yokota O, et al. Positive affect and regional cerebral blood flow in Alzheimer's disease. Psychiatry Res Neuroimaging 2016 Oct 30;256:15-20. [doi: 10.1016/j.pscychresns.2016.09.003] [Medline: 27640073]
15. Liu C, Yi X, Li T, Xu L, Hu M, Zhang S, et al. Associations of depression, anxiety and PTSD with neurological disability and cognitive impairment in survivors of moyamoya disease. Psychol Health Med 2019 Jan;24(1):43-50. [doi: 10.1080/13548506.2018.1467024] [Medline: 29678126]
16. Stanton K, Watson D. Positive and negative affective dysfunction in psychopathology. Soc and Personal Psychol Compass 2014;8(9):555-567. [doi: 10.1111/spc3.12132]
17. Watson D, Vaidya JG. Mood measurement: current status and future directions. In: Schinka JA, Velicer WF, editors. Handbook of Psychology, Volume 2: Research Methods in Psychology. Hoboken, NJ: John Wiley & Sons; 2003:351-375.
18. Everson-Rose SA, Lewis TT. Psychosocial factors and cardiovascular diseases. Annu Rev Public Health 2005;26:469-500. [doi: 10.1146/annurev.publhealth.26.021304.144542] [Medline: 15760298]
19. Brotman DJ, Golden SH, Wittstein IS. The cardiovascular toll of stress. Lancet 2007 Sep 22;370(9592):1089-1100. [doi: 10.1016/S0140-6736(07)61305-1] [Medline: 17822755]
20. Hoffman L. Longitudinal Analysis: Modeling Within-Person Fluctuation and Change. New York: Taylor & Francis; 2015.
21. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002;24(3):236-243. [doi: 10.1207/S15324796ABM2403_09] [Medline: 12173681]
22. Augustine AA, Larsen RJ. Emotion research. In: Mehl MR, Conner TS, editors. Handbook of Research Methods for Studying Daily Life. New York: The Guilford Press; 2012:497-510.
23. Kim J, Lim S, Min YH, Shin Y, Lee B, Sohn G, et al. Depression screening using daily mental-health ratings from a smartphone application for breast cancer patients. J Med Internet Res 2016 Aug 4;18(8):e216 [FREE Full text] [doi: 10.2196/jmir.5598] [Medline: 27492880]
24. Schembre SM, Liao Y, O'Connor SG, Hingle MD, Shen S, Hamoy KG, et al. Mobile ecological momentary diet assessment methods for behavioral research: systematic review. JMIR Mhealth Uhealth 2018 Nov 20;6(11):e11170 [FREE Full text] [doi: 10.2196/11170] [Medline: 30459148]
25. Doherty ST, Lemieux CJ, Canally C. Tracking human activity and well-being in natural environments using wearable sensors and experience sampling. Soc Sci Med 2014 Apr;106:83-92. [doi: 10.1016/j.socscimed.2014.01.048] [Medline: 24549253]
26. Yang YS, Ryu GW, Choi M. Methodological strategies for ecological momentary assessment to evaluate mood and stress in adult patients using mobile phones: systematic review. JMIR Mhealth Uhealth 2019 Apr 1;7(4):e11215 [FREE Full text] [doi: 10.2196/11215] [Medline: 30932866]
27. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009 Sep;39(9):1533-1547. [doi: 10.1017/S0033291708004947] [Medline: 19215626]

28. Farmer S, Mindry D, Comulada WS, Swendeman D. Mobile phone ecological momentary assessment of daily stressors among people living with HIV: Elucidating factors underlying health-related challenges in daily routines. J Assoc Nurses AIDS Care 2017;28(5):737-751 [FREE Full text] [doi: 10.1016/j.jana.2017.04.001] [Medline: 28549526]
29. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
30. Wichers M, Simons CJ, Kramer IM, Hartmann JA, Lothmann C, Myin-Germeys I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011 Oct;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
31. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
32. Yang YS, Ryu GW, Han I, Oh S, Choi M. Ecological momentary assessment using smartphone-based mobile application for affect and stress assessment. Healthc Inform Res 2018 Oct;24(4):381-386 [FREE Full text] [doi: 10.4258/hir.2018.24.4.381] [Medline: 30443427]
33. Kang Y, Na DL, Hahn S. A validity study on the korean mini-mental state examination (K-MMSE) in dementia patients. J Korean Neurol Assoc 1997;15(2):300-308 [FREE Full text]
34. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
35. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
36. Lee E, Chung BY, Suh C, Jung J. Korean versions of the Perceived Stress Scale (PSS-14, 10 and 4): psychometric evaluation in patients with chronic disease. Scand J Caring Sci 2015 Mar;29(1):183-192. [doi: 10.1111/scs.12131] [Medline: 24660854]
37. Verhagen SJ, Berben JA, Leue C, Marsman A, Delespaul PA, van Os J, et al. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. PLoS One 2017;12(10):e0186294 [FREE Full text] [doi: 10.1371/journal.pone.0186294] [Medline: 29023568]
38. Powell DJ, Schlotz W. Daily life stress and the cortisol awakening response: testing the anticipation hypothesis. PLoS One 2012;7(12):e52067 [FREE Full text] [doi: 10.1371/journal.pone.0052067] [Medline: 23284871]
39. Petrowski K, Paul S, Albani C, Brähler E. Factor structure and psychometric properties of the trier inventory for chronic stress (TICS) in a representative German sample. BMC Med Res Methodol 2012 Apr 1;12:42 [FREE Full text] [doi: 10.1186/1471-2288-12-42] [Medline: 22463771]
40. Pala A, Pawlikowski A, Brand C, Schmitz B, Wirtz CR, König R, et al. Quality of life after treatment of unruptured intracranial aneurysms. World Neurosurg 2019 Jan;121:e54-e59. [doi: 10.1016/j.wneu.2018.09.010] [Medline: 30244183]
41. Arwert HJ, Meesters JJ, Boiten J, Balk F, Wolterbeek R, Vlieland TP. Poststroke depression: a long-term problem for stroke survivors. Am J Phys Med Rehabil 2018 Aug;97(8):565-571. [doi: 10.1097/PHM.0000000000000918] [Medline: 29509550]
42. Laures-Gore JS, Defife LC. Perceived stress and depression in left and right hemisphere post-stroke patients. Neuropsychol Rehabil 2013;23(6):783-797 [FREE Full text] [doi: 10.1080/09602011.2013.811087] [Medline: 23799820]
43. Guo L, Liu Y, McCallum J, Söderhamn U, Ding X, Yv S, et al. Perceived stress and depression amongst older stroke patients: Sense of coherence as a mediator? Arch Gerontol Geriatr 2018;79:164-170. [doi: 10.1016/j.archger.2018.08.010] [Medline: 30265911]
44. Zhang Y, Tuomilehto J, Jousilahti P, Wang Y, Antikainen R, Hu G. Lifestyle factors on the risks of ischemic and hemorrhagic stroke. Arch Intern Med 2011 Nov 14;171(20):1811-1818. [doi: 10.1001/archinternmed.2011.443] [Medline: 21911621]
45. Lee MS, Park CG, Hughes TL, Jun SE, Whang K, Kim N. The predictive role of health-promoting behaviours and perceived stress in aneurysmal rupture. J Clin Nurs 2018 Mar;27(5-6):e1068-e1077. [doi: 10.1111/jocn.14149] [Medline: 29076614]
46. Ditzen B, Heinrichs M. Psychobiology of social support: the social dimension of stress buffering. Restor Neurol Neurosci 2014;32(1):149-162. [doi: 10.3233/RNN-139008] [Medline: 23603443]
47. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. Psychol Bull 1985 Sep;98(2):310-357. [Medline: 3901065]


## Abbreviations

**EMA:** ecological momentary assessment
**K-HADS:** Korean version of the Hospital Anxiety and Depression Scale
**MMD:** Moyamoya disease
**MMS:** Moyamoya syndrome
**MR:** magnetic resonance
**NA:** negative affect
**PA:** positive affect

*Edited by G Eysenbach; submitted 13.11.19; peer-reviewed by K Aguirre, D Saleh; comments to author 04.01.20; revised version received 29.01.20; accepted 06.02.20; published 25.05.20.*

*Please cite as:*
*Yang YS, Ryu GW, Park CG, Yeom I, Shim KW, Choi M*
*Mood and Stress Evaluation of Adult Patients With Moyamoya Disease in Korea: Ecological Momentary Assessment Method Using a Mobile Phone App*
*JMIR Mhealth Uhealth 2020;8(5):e17034*
*URL: http://mhealth.jmir.org/2020/5/e17034/*
*doi:10.2196/17034*
*PMID:32449687*

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org