JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 14 (2026) ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoint

Reviews

Original Papers

Tutorial

Viewpoint

Intensive, Repeated Self-Report Measures: Should We Be Concerned About Changes in Data Quality Over Time?

Arthur A Stone[1*], PhD; Stefan Schneider[1*], PhD; Meynard J Toledo[1*], PhD; Raymond Hernandez[1], PhD

Department of Psychology, University of Southern California, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Arthur A Stone, PhD
Department of Psychology
University of Southern California
McClintock Street
Los Angeles, CA, 90290
United States
Phone: 1 6316820014
Email: arthur.stone@usc.edu

Abstract

Intensive, repeated self-report measures are an important tool for behavioral and medical researchers and practitioners who are concerned with the dynamic interplay among variables at a granular level. Many mobile health applications rely on accurate measurement of immediate states and environments for both assessment and intervention delivery. Techniques for capturing repeated momentary assessments yield data with several salutary qualities: recall bias is minimized relative to assessments that rely on much longer recall periods; measurements are taken in individuals' everyday environments; and dense, repeated measures allow a new window into the processes transpiring between individuals and their environments. In this paper, we highlight several features of repeatedly completing momentary assessments that may change the nature or quality of the data collected over time. Several lines of inquiry are discussed that call into question the presumption that there is invariance in how people complete repeated assessments over time. A result of this possibility could be a reduction in data quality. We present 4 phenomena, with selected results, that may induce noninvariance in repeated measures: the amount of time required to complete assessments, the rate of missing data, the degree of careless responding, and the presence of several components of reactivity. In each of these areas, we found evidence that changes could occur over time, and we consider how data might be affected by such changes. Our conclusion is that researchers should be aware that changes can occur over time and that these changes may affect data quality.

(JMIR Mhealth Uhealth 2026;14:e68735) doi:10.2196/68735

KEYWORDS

ecological momentary assessment; EMA; repeated measurements; invariance; self-report

Introduction

Repeated self-report measures are central to intensive longitudinal methods that use momentary assessments, including ecological momentary assessment (EMA) [1,2], the experience sampling method [3], and ambulatory assessment methods [4]. Methodological strategies derived from momentary assessment studies form the backbone of optimal mobile health (mHealth) data acquisition for participant monitoring, in-the-moment treatment delivery (eg, just-in-time adaptive interventions), and real-time feedback on behavior and thoughts. As such, there is considerable value in recognizing the strengths and the potential limitations of these methods.

The advantages of repeated measures over static measures are numerous: they provide the opportunity to track trajectories of outcomes over time, enable exploration of lagged associations that may confer near-causal status, and reduce problems associated with retrospection [2,5-7]. Repeated-measures study designs are also part and parcel of many mHealth practices. Moreover, constructs often measured using momentary methods are pertinent for mHealth researchers; they include internal states (eg, pain intensity and quality, emotions, fatigue, perceived stress, symptoms, and cognitive status) and external (often observable) behaviors (eg, daily activities, social engagement, consumption, substance use, exogenous events, and location). Real-time self-reports may also be linked to ambulatory measures of physiological function (eg, heart rate, blood pressure, cortisol levels, and blood glucose) and to data

routinely collected by smartphones (eg, location, step counts, and time taken to respond to questions) to provide insights into the dynamics of these variables. Finally, intensive, repeated measures are indispensable tools for monitoring populations, evaluating treatments in medical research, analyzing economic patterns, and addressing behavioral science questions. Many review and position papers [6,8-11] are available for the interested reader.

A generally unspoken belief about repeated self-report measures is that repeated assessments conducted throughout a study—regardless of the study duration or the momentary measurement frequency—are thought to be invariant regarding how questions are interpreted, how response scales are construed, and how internal reference standards are applied to ratings. All of these are important considerations for achieving reliable and valid data. That is, we assume that the quality and integrity of responses to all assessments are not affected by repeated measurement processes. However, if this assumption of invariance does not hold, it becomes challenging to disentangle true within-person changes from shifts in measurement quality over time. For example, observed within-person patterns could be erroneously attributed to genuine change when they instead reflect evolving interpretations of questions or changes in response scale use, thereby compromising the validity of inferences about within-person dynamics and processes.

Our primary intention in this viewpoint is to highlight several features of repeated momentary assessments that we think deserve attention, given the possibility that they are associated with changes in momentary data quality. We present selected results to illustrate the point. Because the findings presented later have been reported previously, it is reasonable to assume that they may be present in at least some repeated momentary studies and, importantly, *may* systematically impact data quality and warrant further research. In the same spirit, we do not argue that the features presented are ubiquitous in EMA studies.

We now describe how intensive longitudinal momentary measurements may be distorted by repeatedly asking the same or similar questions over many repetitions. Theoretically, these distortions may be particularly germane when assessments occur many times a day, with only hours between assessments [12], but they may also occur at longer intervals. There is already substantial evidence that responses are not invariant over time, given that repeated measures can create practice effects, a phenomenon supported by abundant research on memory and performance [13-15]. Practice effects could be salutary, leading to faster and more accurate responses [16], or pernicious, leading to unwanted measurement effects such as increased bias and error [17]. In either case, the key point is that repeated exposure to the content can change how questions are answered. To date, evidence on the *adverse* consequences of repeated assessments as they pertain to EMA has been available in a piecemeal fashion spanning several content domains, making it difficult to fully appreciate the scope and importance of the issue.

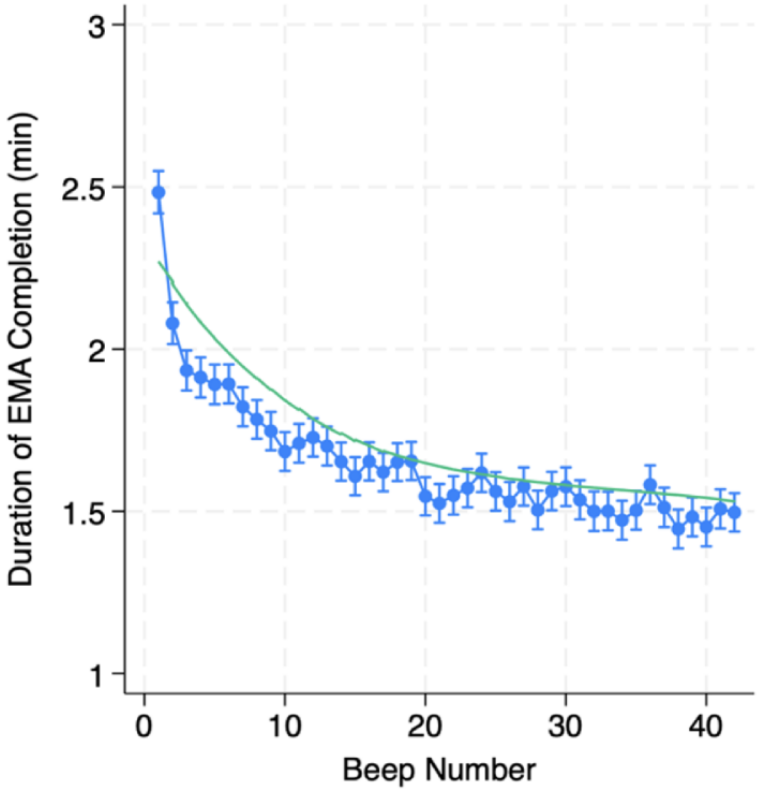
Response Times Decrease for EMA Question Completion Over Repeated Measures

Overview

Although perhaps not well known to EMA researchers, studies examining the time taken to complete an EMA assessment have reported decreases in response time (time it takes individuals to complete questions) over repeated measures, consistent with the practice effects mentioned previously [18-20]. Response time reductions are likely due to respondents' increasing familiarity with the task, with the wording and meaning of questions, and with the response options over repeated assessments [21]. These decrements can be striking, with reductions of more than 50%. Aggregated response times usually follow a negative exponential pattern, with the steepest decrements occurring early during repeated measures.

To illustrate this pattern of response times over time in an EMA study, data from the Understanding American Study conducted at University of Southern California are shown in Figure 1. A total of 22,531 prompts were recorded from 706 community-dwelling adults, and up to 42 prompts were completed per week. Average response time is plotted for each prompt, revealing a dramatic drop across the first 9 prompts followed by a modest decline over the remaining prompts. Over the course of a week, there was a 32% decline in response times. The magnitude of the decrease in survey completion time is influenced by multiple factors, including the burden of the interview, incentives for completion, and the composition of the participant sample.

Figure 1. Response times over prompts. EMA: ecological momentary assessment.



Implications

We conclude that the time to complete assessments may decrease over repeated measures. We speculate that if such changes are associated with, or caused by, any of the factors discussed later, data quality may change over repeated measures. For example, increasing missingness and carelessness over time may generate faster response times and, consequently, compromise data quality. Given this possibility, we suggest that researchers examine the response time changes over the course of repeated measures, though we also acknowledge that response time may or may not be associated with validity.

Rate of Data Missingness May Increase Over Repeated Measures

Overview

One plausible explanation for faster survey completion with repeated measures is that individuals gradually disengage from the study procedures and, therefore, spend less time completing assessments. At the start of a study, participants may be motivated to complete all questions and generate optimal answers but may lose interest and become less invested as they repeatedly answer the same questions [22]. When this happens, participants may show increasing noncompliance with the study protocol and may miss assessments when prompted, a concern that has previously been acknowledged by EMA researchers [23].

To assess the magnitude of change in missed EMA prompts over time, we examined many EMA studies with a wide range of samples and variety of assessment schedules. Some studies showed a drop in compliance rates over time, across both shorter and longer studies, and with various frequencies of EMA prompting. Therefore, we suggest that repeated measures can induce *some level* of missingness in EMA studies.

We now turn to the question of whether changes in missing data are serious contenders for inducing bias and how such missingness may affect data quality. If shifts in missingness over time are minimal or if compliance rates remain high overall despite an increase in missed assessments, there is likely little cause of concern. However, in the presence of substantial increases in missingness, missing data can introduce a range of problems, including reduced statistical power, unrepresentative momentary samples (or changes in sample representation), and potential distortions of longitudinal effects. Increasing rates of missing values imply that missing values do not occur completely at random, and researchers may wish to routinely incorporate study day (or prompt number) as an auxiliary variable or covariate in statistical analyses to reduce potential bias due to systematic missingness patterns (ie, to increase the plausibility of the assumption that data are missing at random. Although statistical methods (eg, multilevel multiple imputation [24]) can help account for missing-at-random mechanisms, it is not possible to fully protect against potential patterns that are missing not at random, and only a few studies have developed or evaluated missing data methods specifically for intensive longitudinal contexts, such as EMA data [25].

Implications

Repeated measures may create bias when missingness increases to substantial levels over the course of a study, and we suggest that missingness over time be considered in intensive repeated-measures studies. Unfortunately, the precise proportion

of missingness that would signal caution depends on the cause of this missingness (eg, simple burden or a change in activities that encourage missingness, such as more exercise). Our recommendation is to implement as many safeguards as possible to encourage the most complete data throughout an entire study. This could entail a variety of methodological maneuvers: creating EMA designs that reduce respondent burden related to questionnaire length, daily sampling frequency and scheduling, and study duration; incentive structures that encourage high compliance (eg, payment being contingent on prespecified completion rules from the investigators); and implementing procedures for monitoring compliance over time and providing appropriate feedback to encourage compliance (eg, real-time or daily signaling to the research team about missed prompts). Further research would be welcome to determine which of these strategies are most effective, so that efforts to reduce missingness could be implemented efficiently.

Frequency of Careless Responding Over Time

Overview

Careless responding (or insufficient effort responding) is defined as respondents providing answers without regard to the content of the questions, and it can occur when they do not read an item or do not pay attention to what the item is asking [26,27]. Careless responding can also occur as a more subtle form of participant disengagement compared with the overt noncompliance discussed previously, in which a respondent misses prompts entirely. To reduce the burden associated with repeated assessments, respondents may minimize the effort expended by doing the least possible to satisfy study requirements [22]. Careless responding can present as invariable responses (eg, *straight-lining*, in which the same answer is given to every question) or as inconsistent or random responses. A recent small study interviewed individuals at the close of a many-month EMA study to better understand the burden and other issues that could affect data quality [28]. The authors found that respondents tried to counter the burden by responding quickly and that repeated measures over time led to more neutral responses.

Many studies have shown the detrimental effects that careless responding can have on measurement accuracy and reliability [26]. Careless responding can occur when respondents are asked to complete long questionnaires. For instance, toward the end of lengthy questionnaires with dozens or hundreds of items, respondents often give more random, uniform, and fast responses, suggesting that their motivation has waned over time [29,30]. An intuitive extrapolation of these findings is that the repetitive nature of completing brief surveys may similarly lead to more careless responding. Ganzach and Bulmash [31] documented less variable and less complex patterns of self-reporting over the course of daily repeated measures on affect and stress. These findings demonstrated that increased carelessness could occur over repeated measures [32]. Whether decreases in survey completion time can be attributed to increases in extremely fast responses (an indicator of carelessness) has also been shown to occur. In another EMA study, the frequency of speed responding increased from 3% to 8% across 49 assessments over 7 days [33]. Again, we are not saying that all studies, or even most studies, show this pattern—only that it can occur in repeated-measures studies. Another method for assessing changes in carelessness over time is by comparing self-report ratings with objective assessments of the same concept. With increased carelessness, associations over repeated measures would be predicted to decrease because of the increased measurement error. One study found a decreasing association between self-reported and objectively measured time spent on social media in adolescents, which could have been generated through carelessness [34].

Implications

Available evidence shows instances of increasing carelessness over densely repeated measures. We recommend future work to examine possible changes in careless responding over time. The addition of attention check questions might also improve engagement and reduce carelessness [28]. A fruitful direction for future research could be the development of advanced statistical models, including machine learning techniques, to detect careless responding in datasets in which EMA self-reports can be compared with objective measures. Furthermore, the longitudinal aspect of intensive repeated-measures data may prove a fertile ground for extending carelessness detection approaches, for instance, by examining shifts in within-person parameters (eg, variability, autocorrelation, and outliers) over time.

Reactivity to Question Content May Increase Over Repeated Measures

Overview

Reactivity, or reactive arrangements, has been recognized since the 1930s, when workers at a manufacturing facility in Hawthorne, New Jersey, altered their behavior in response to their actions being observed by scientists [35], even though an alternative interpretation has been offered [36]. Here, we restrict our inquiry to reactivity associated with repeated measures in intensive momentary designs, namely, the repeated answering of the same or similar questions many times. This section is divided into 3 subparts: reactivity producing a change in a targeted behavior or cognition, reactivity producing a change in the reporting of a behavior or cognition (but without corresponding change in the targeted phenomenon), and specific reactivity processes associated with changes in question meaning or scale recalibration. These are not exclusive designations, and it is possible that studies exhibit blends of these types of reactivity.

Reactivity in Behavior, Cognition, and Emotion

This type of reactivity is defined as a change in actual behavior, cognition, and emotion over time due to repeated responses. In fact, some interventions have intentionally incorporated this type of reactivity to induce desired outcomes, such as the use of self-monitoring to affect physical activity behaviors [37,38]. Despite its potential utility in intervention contexts, this phenomenon poses a threat to the internal validity of an observational study as it signifies that measurement is not

invariant over time. For example, repeated assessments may cause participants to become more conscious of their behaviors, which may, in turn, trigger self-regulatory mechanisms [39]. The same process may happen with behaviors and emotions. Unfortunately, very few studies are available on whether and how this phenomenon unfolds in EMA studies over time (eg, potential for cumulative effects) [40].

Reactivity in Reporting

A second type of reactivity is defined by respondents altering the way they *report* behaviors, cognitions, and emotions, independent of whether those behaviors or states have actually changed over time (the previous type of reactivity). In repeated-measures studies, this phenomenon could manifest as a shift in the validity of the reported variables over repeated assessments. For example, reports of affect might be accurate at the onset of a study but gradually shift to become inaccurate or less reliable later; the converse might also be true, as discussed later [18]. A problem for reporting reactivity is that most studies cannot distinguish between real change and reporting change because doing so requires objective measurement of the variable of interest, which is very uncommon in the constructs being measured in EMA. This requirement presents a challenge for the assessment of internal states, where changes over time may indicate actual change, reporting change, or a combination of both.

Reactivity Processes

The third section on reactivity discusses several ways that actual levels of, or reports of, behavior, cognition, and emotion might change over time through specific psychological processes that may be inherent in repeated measures.

Initial elevation is defined as higher levels of a variable when measured early in a study compared with levels observed at later measurements when there is no apparent reason for a shift in levels over time [41-43]. Although recognized for many years, a paper by Shrout et al [43] has renewed interest in the phenomenon. They found an initial elevation effect for momentary negative states and for momentary internal states in 4 studies with repeated measures, with small to medium effect sizes. Furthermore, Anvari et al [44] (2023) studied several thousand college students and reported a strong initial elevation bias. Thus, initial elevation has been observed, although there is some competing evidence [45] consistently. Biasing of overall levels of variables can result from such effects, for example, by producing (apparently) incorrect downward trajectories in variables over time.

Another facet of reactivity that has not received much attention is that the *meaning of questions* may evolve over repeated administrations of questions [46]. Following up on previous work by Windle [47] (1994), Knowles et al [48] (1996) explicitly explored the possibility that the interpretation of questions changes over time. Results indicated that over multiple exposures to items, individuals gained knowledge about the construct being studied (an anxiety questionnaire). Moreover, they observed shifts in question meaning within a single test administration. Later interpretations of anxiety items demonstrated that a higher standard for endorsing anxiety was evident; if all else remained constant, this would result in diminished reported anxiety later in the study. The upshot of such shifts in question meaning over repeated measures over time is clear: what we learn at the beginning of a study might not be comparable with what we learn later.

Another aspect of question meaning is *responsescale recalibration*, defined as a shift in how participants use question response options that occurs as a reaction to events experienced throughout the study [46]. Exposures to events that elicit very intense emotions or other strong reactions (eg, pain of childbirth or trauma) may alter the way individuals interpret a rating scale. Such recalibration could result in a stimulus that was once rated at a certain intensity to later receive a lower intensity rating because the upper end of the scale was redefined by the extreme emotion or reaction. Alternatively, recalibration could be induced in certain treatment modalities intended to reduce reported symptomatology through cognitive and social manipulations [49]. Therefore, recalibration could have a major impact on interpreting results over time.

Implications

It is fundamental to know if repeatedly measuring a variable creates any of the reactivity processes, given that they would be a threat to internal validity. If there is a high likelihood for reactivity, we recommend preliminary investigations to assess the potential magnitude of these effects and to incorporate design strategies to mitigate them, while acknowledging the extra financial and investigator burden inherent in this recommendation. More broadly, if empirical evidence suggests that reactivity is likely (eg, respondents having access to feedback that has been previously shown to induce reactivity), steps could be taken to eliminate the reactivity-inducing aspect of the protocol.

An existing barrier to the application of this approach is that we simply do not have an adequate way to identify reactivity-inducing circumstances. Of particular importance is the possibility that processes and events with the potential of shifting means or recalibrating scales are not routinely measured in longitudinal studies; we therefore recommend routine collection of such data to signal possible distortion. In addition, experimental manipulations of item administration (eg, item order) have been recommended as useful methods to detect possible biases from measurement reactivity [19]. Advancing methodologies to statistically control for these shifts is also recommended; however, some efforts to accomplish this have not resulted in reduced bias, such as using the *then-test* to control for shifts from pretest to posttest [49].

Conclusions

The aim of this viewpoint was to consider the possibility that studies using intensive repeated self-report measures could distort interpretations of data collected over time. We found evidence supporting the potential for all factors considered: response times can be considerably reduced, missing data rates can increase, carelessness rates can rise, and reactive effects (including reference standards and question meaning) may manifest over repeated measures. Response time reductions

alone may or may not impact data quality, as practice effects could be positive or negative. However, when reduced response times lead to increases in missingness, carelessness, and reactivity, there is the potential for compromised data quality. Heightened awareness of these possibilities is warranted for researchers using repeated assessments. We also offered suggestions for monitoring and reducing these possible effects.

We view this discussion as a supplement to prior papers that suggested *guidelines* for conducting, reporting, and analyzing EMA studies. Those papers have guidelines that could be helpful for examining some threats to validity across repeated measures, although not with an eye toward invariance over time. As early as 2002, an EMA guidelines publication [50] indicated the need for comprehensive reporting of procedures and enumerated ways of calculating prompt compliance, which are useful for understanding the threats to validity discussed here. The early set of guidelines did not mention tracking the duration of prompt completion or methods for assessing careless responding. A guidance document [51] included the recommendation to record prompt duration; these data would be pertinent for evaluating the issues raised in this paper. However, their suggestions did not mention tracking careless responding. Overall, these recommendations are positive developments for the field that should result in more robust EMA studies.

In summary, we believe a cautious position is warranted regarding the possibility that repeated questionnaire measurements affect how questions are answered. It is plausible that some results from repeated-measures studies have been affected by the processes discussed previously. This paper calls for research directed at understanding whether reactivity distorts data collected through intensive repeated measures. A speculative extension of this conclusion concerns repeated measures taken at longer intervals, such as every week or month: such longitudinal data may also be susceptible to the biases reported here, with potentially far-reaching implications.

Funding

This study was supported by award R37AG057685 (AAS, principal investigator).

Conflicts of Interest

AAS is a senior scientist with the Gallup Organization and a consultant for Lore Contagious, Inc and AstraZeneca, Inc. All other authors declare no other conflicts of interest.

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. Ann Behav Med 1994;16(3):199-202. [doi: 10.1093/abm/16.3.199]
3. deVries MW. Investigating mental disorders in their natural settings: introduction to the special issue. J Nerv Ment Dis 1987 Sep;175(9):509-513. [doi: 10.1097/00005053-198709000-00001] [Medline: 3655774]
4. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: a review of recommended reporting guidelines and current practices. J Abnorm Psychol 2020 Jan;129(1):56-63. [doi: 10.1037/abn0000473] [Medline: 31868388]
5. Stone AA, Turkkan JS, Jobe JB, Bachrach C, Kurtzman HS, Cain VS. The Science of Self-Report: Implications for Research and Practice. London, UK: Lawrence Erlbaum Associates; 2000.
6. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York, NY: Guilford Press; 2013.
7. 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]
8. Smyth JM, Stone AA. Ecological momentary assessment research in behavioral medicine. J Happiness Stud 2003 Mar;4:35-52. [doi: 10.1023/a:1023657221954]
9. Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. Psychol Assess 2009 Dec;21(4):463-475. [doi: 10.1037/a0017075] [Medline: 19947781]
10. Stone AA, Broderick JE, Schneider S, Schwartz JE. Expanding options for developing outcome measures from momentary assessment data. Psychosom Med 2012 May;74(4):387-397 [FREE Full text] [doi: 10.1097/PSY.0b013e3182571faa] [Medline: 22582336]
11. Stone AA, Obbarius A, Junghaenel DU, Wen CK, Schneider S. High-resolution, field approaches for assessing pain: ecological momentary assessment. Pain 2021 Jan;162(1):4-9 [FREE Full text] [doi: 10.1097/j.pain.0000000000002049] [Medline: 32833794]
12. Robinson MD, Clore GL. Belief and feeling: evidence for an accessibility model of emotional self-report. Psychol Bull 2002;128(6):934-960. [doi: 10.1037//0033-2909.128.6.934]

13. Bartels C, Wegrzyn M, Wiedl A, Ackermann V, Ehrenreich H. Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. BMC Neurosci 2010 Sep 16;11:118 [FREE Full text] [doi: 10.1186/1471-2202-11-118] [Medline: 20846444]
14. Holm SP, Wolfer AM, Pointeau GH, Lipsmeier F, Lindemann M. Practice effects in performance outcome measures in patients living with neurologic disorders - a systematic review. Heliyon 2022 Aug;8(8):e10259 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e10259] [Medline: 36082322]
15. Duff K, Callister C, Dennett K, Tometich D. Practice effects: a unique cognitive variable. Clin Neuropsychol 2012;26(7):1117-1127. [doi: 10.1080/13854046.2012.722685] [Medline: 23020261]
16. van den Broek GS, Segers E, Takashima A, Verhoeven L. Do testing effects change over time? Insights from immediate and delayed retrieval speed. Memory 2014 Sep 02;22(7):803-812. [doi: 10.1080/09658211.2013.831455] [Medline: 23998337]
17. Schnell PM, Wascher M, Rempala GA. Overcoming repeated testing schedule bias in estimates of disease prevalence. J Am Stat Assoc 2023 Sep 06;119(545):1-13. [doi: 10.1080/01621459.2023.2238943]
18. Hernandez R, Schneider S, Pinkham AE, Depp CA, Ackerman R, Pyatak EA, et al. Comparisons of self-report with objective measurements suggest faster responding but little change in response quality over time in ecological momentary assessment studies. Assessment 2025 Apr;32(3):335-355. [doi: 10.1177/10731911241245793] [Medline: 38634454]
19. Arslan RC, Reitz AK, Driebe JC, Gerlach TM, Penke L. Routinely randomize potential sources of measurement reactivity to estimate and adjust for biases in subjective reports. Psychol Methods 2021 Apr;26(2):175-185. [doi: 10.1037/met0000294] [Medline: 32584065]
20. Schneider S, Hernandez R, Junghaenel DU, Orriens B, Lee PJ, Stone AA. Response times in ecological momentary assessment (EMA): shedding light on the response process with a drift diffusion model. Curr Psychol 2024 Feb 27;43(7):5868-5886. [doi: 10.1007/s12144-023-04773-0] [Medline: 40969298]
21. McCabe D, Langer KG, Borod JC, Bender HA. Practice effects. In: Kreutzer JS, DeLuca J, Caplan B, editors. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer; 2011.
22. Krosnick JA. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl Cognit Psychol 2006 Feb 13;5(3):213-236. [doi: 10.1002/acp.2350050305]
23. Wrzus C, Neubauer AB. Ecological momentary assessment: a meta-analysis on designs, samples, and compliance across research fields. Assessment 2023 Apr 11;30(3):825-846 [FREE Full text] [doi: 10.1177/10731911211067538] [Medline: 35016567]
24. Ji L, Li Y, Potter LN, Lam CY, Nahum-Shani I, Wetter DW, et al. Multiple imputation of missing data in multilevel ecological momentary assessments: an example using smoking cessation study data. Front Digit Health 2023 Nov 10;5:1099517 [FREE Full text] [doi: 10.3389/fdgth.2023.1099517] [Medline: 38026834]
25. Goldberg SB, Bolt DM, Davidson RJ. Data missing not at random in mobile health research: assessment of the problem and a case for sensitivity analyses. J Med Internet Res 2021 Jun 15;23(6):e26749 [FREE Full text] [doi: 10.2196/26749] [Medline: 34128810]
26. Ward MK, Meade AW. Dealing with careless responding in survey data: prevention, identification, and recommended best practices. Annu Rev Psychol 2023 Jan 18;74(1):577-596 [FREE Full text] [doi: 10.1146/annurev-psych-040422-045007] [Medline: 35973734]
27. Huang JL, Liu M, Bowling NA. Insufficient effort responding: examining an insidious confound in survey data. J Appl Psychol 2015 May;100(3):828-845. [doi: 10.1037/a0038510] [Medline: 25495093]
28. Wang SD, Hatzinger L, Morales J, Hewus M, Intille S, Dunton GF. Burden and inattentive responding in a 12-month intensive longitudinal study: interview study among young adults. JMIR Form Res 2024 Aug 02;8:e52165 [FREE Full text] [doi: 10.2196/52165] [Medline: 39093606]
29. Bowling NA, Gibson AM, Houpt JW, Brower CK. Will the questions ever end? Person-level increases in careless responding during questionnaire completion. Organ Res Methods 2020 Aug 26;24(4):718-738. [doi: 10.1177/1094428120947794]
30. Galesic M, Bosnjak M. Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opin Q 2009 Jun;73(2):349-360. [doi: 10.1093/poq/nfp031]
31. Ganzach Y, Bulmash B. The effect of serial day on the measurement of positivity and emotional complexity in diary studies. Eur J Soc Psychol 2021 Dec;51(7):1213-1225. [doi: 10.1002/ejsp.2809]
32. Gochmann V, Ohly S, Kotte S. Diary studies, a double-edged sword? An experimental exploration of possible distortions due to daily reporting of social interactions. J Organ Behav 2022 Sep;43(7):1209-1223. [doi: 10.1002/job.2633]
33. Ulitzsch E, Nestler S, Lüdtke O, Nagy G. A screen-time-based mixture model for identifying and monitoring careless and insufficient effort responding in ecological momentary assessment data. Psychol Methods 2024 Feb 29. [doi: 10.1037/met0000636] [Medline: 38421768]
34. Verbeij T, Pouwels JL, Beyens I, Valkenburg PM. The accuracy and validity of self-reported social media use measures among adolescents. Comput Hum Behav Rep 2021;3:100090. [doi: 10.1016/j.chbr.2021.100090]
35. Gillespie R. Manufacturing Knowledge: A History of the Hawthorne Experiments. Cambridge, UK: Cambridge University Press; 1993.

36. Adair JG. The Hawthorne effect: a reconsideration of the methodological artifact. J Appl Psychol 1984 May;69(2):334-345. [doi: 10.1037/0021-9010.69.2.334]
37. Compernolle S, DeSmet A, Poppe L, Crombez G, De Bourdeaudhuij I, Cardon G, et al. Effectiveness of interventions using self-monitoring to reduce sedentary behavior in adults: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2019 Aug 13;16(1):63 [FREE Full text] [doi: 10.1186/s12966-019-0824-3] [Medline: 31409357]
38. Kanejima Y, Kitamura M, Izawa KP. Self-monitoring to increase physical activity in patients with cardiovascular disease: a systematic review and meta-analysis. Aging Clin Exp Res 2019 Feb;31(2):163-173. [doi: 10.1007/s40520-018-0960-7] [Medline: 29714027]
39. Maher JP, Arigo D, Baga K, Salvatore GM, Pasko K, Hudgins B, et al. Measurement reactivity in ecological momentary assessment studies of movement-related behaviors. J Meas Phys Behav 2024 Jan;7(1) [FREE Full text] [doi: 10.1123/jmpb.2023-0035]
40. Eisele G, Vachon H, Lafit G, Tuyaerts D, Houben M, Kuppens P, et al. A mixed-method investigation into measurement reactivity to the experience sampling method: the role of sampling protocol and individual characteristics. Psychol Assess 2023 Jan;35(1):68-81. [doi: 10.1037/pas0001177] [Medline: 36174163]
41. Johar O, Sackett AM. The self-contaminating nature of repeated reports of negative emotions. Basic Appl Soc Psychol 2018;40(5):293-307. [doi: 10.1080/01973533.2018.1496336] [Medline: 38634454]
42. French DP, Sutton S. Reactivity of measurement in health psychology: how much of a problem is it? What can be done about it? Br J Health Psychol 2010 Sep;15(Pt 3):453-468. [doi: 10.1348/135910710X492341] [Medline: 20205982]
43. Shrout PE, Stadler G, Lane SP, McClure MJ, Jackson GL, Clavél FD, et al. Initial elevation bias in subjective reports. Proc Natl Acad Sci U S A 2018 Jan 02;115(1):E15-E23 [FREE Full text] [doi: 10.1073/pnas.1712277115] [Medline: 29255039]
44. Anvari F, Efendić E, Olsen J, Arslan RC, Elson M, Schneider IK. Bias in self-reports: an initial elevation phenomenon. Soc Psychol Pers Sci 2022 Oct 07;14(6):727-737. [doi: 10.1177/19485506221129160]
45. Cerino ES, Schneider S, Stone AA, Sliwinski MJ, Mogle J, Smyth JM. Little evidence for consistent initial elevation bias in self-reported momentary affect: a coordinated analysis of ecological momentary assessment studies. Psychol Assess 2022 May;34(5):467-482 [FREE Full text] [doi: 10.1037/pas0001108] [Medline: 35175074]
46. Ubel PA, Peeters Y, Smith D. Abandoning the language of "response shift": a plea for conceptual clarity in distinguishing scale recalibration from true changes in quality of life. Qual Life Res 2010 May 29;19(4):465-471. [doi: 10.1007/s11136-010-9592-x] [Medline: 20112000]
47. Windle C. Test-retest effect on personality questionnaires. Educ Psychol Meas 1954 Dec 01;14(4):617-633. [doi: 10.1177/001316445401400404]
48. Knowles ES, Coker MC, Scott RA, Cook DA, Neville JW. Measurement-induced improvement in anxiety: mean shifts with repeated assessment. J Pers Soc Psychol 1996 Aug;71(2):352-363. [doi: 10.1037//0022-3514.71.2.352] [Medline: 8765486]
49. Nolte S, Elsworth GR, Sinclair AJ, Osborne RH. The inclusion of 'then-test' questions in post-test questionnaires alters post-test responses: a randomized study of bias in health program evaluation. Qual Life Res 2012 Apr 28;21(3):487-494. [doi: 10.1007/s11136-011-9952-1] [Medline: 21710355]
50. Stone AA, Shiffman S. Capturing momentary, self-report data: a proposal for reporting guidelines. Ann Behav Med 2002 Aug;24(3):236-243. [doi: 10.1207/s15324796abm2403_09]
51. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021 Nov 19;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]

Abbreviations

EMA: ecological momentary assessment
mHealth: mobile health

Edited by L Buis; submitted 12.Nov.2024; peer-reviewed by A Jamal, C Baxter; comments to author 24.Apr.2025; revised version received 17.Oct.2025; accepted 16.Dec.2025; published 14.Jan.2026.

Please cite as:
Stone AA, Schneider S, Toledo MJ, Hernandez R
Intensive, Repeated Self-Report Measures: Should We Be Concerned About Changes in Data Quality Over Time?
JMIR Mhealth Uhealth 2026;14:e68735
URL: https://mhealth.jmir.org/2026/1/e68735
doi:10.2196/68735
PMID:

Exploiting Unsupervised Free-Living Data for Cardiorespiratory Fitness Estimation: Systematic Review and Meta-Analysis

Alexios Dosis[1,2], MSc, MD; Aron Berger Syversen[3], BSc, MSc; Mikolaj R Kowal[1,2], MBChB, PgDIP; Daniel Grant[4], BSc; Jim Tiernan[2], MBBS, PhD; David Wong[5], MEng, DPhil; David G Jayne[1,2], BSc, MBBCHIR, MD

[1]Leeds Institute of Medical Research, University of Leeds, Clinical Sciences Building, St James's University Hospital, Beckett Street, Leeds, United Kingdom

[2]Abdominal Medicine and Surgery, Leeds Teaching Hospitals, Leeds, United Kingdom

[3]School of Computing, University of Leeds, Leeds, England, United Kingdom

[4]Cardiorespiratory Department, Leeds Teaching Hospitals, Leeds, United Kingdom

[5]5 Leeds Institute of Health Sciences, University of Leeds, Leeds, England, United Kingdom

Corresponding Author:
Alexios Dosis, MSc, MD
Leeds Institute of Medical Research, University of Leeds, Clinical Sciences Building, St James's University Hospital, Beckett Street, Leeds, United Kingdom

Abstract

Background: Current methods of cardiorespiratory fitness (CRF) assessment may discriminate against frail individuals who are challenged to perform a maximal cardiopulmonary exercise test. CRF estimations from free-living wearable data, captured over extended time periods, may offer a more representative assessment and increase usability in clinical settings.

Objective: This study aimed to review current evidence behind this novel concept and evaluate the performance and quality of models developed to estimate CRF from free-living, unsupervised data.

Methods: Following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, we systematically searched 4 databases (MEDLINE, Embase, Scopus, and arXiv) for studies reporting the development of models to estimate CRF from continuous free-living wearable data. Studies conducted entirely under controlled laboratory conditions were excluded. Performance metrics were combined in a meta-correlation analysis using a random-effects model and Fisher Z transformation.

Results: Of 1848 papers screened, 18 met the eligibility criteria, with a total of 31,072 participants. The weighted mean age was 46.9 (SD 1.46) years. Multiple computational techniques were used, with 8 studies employing more advanced machine learning models. The meta-correlation analysis revealed a pooled overall estimate of 0.83 with a 95% CI 0.77 - 0.88. The I^2 test indicated high heterogeneity at 97%. Risk of bias assessment found most concerns in the data analysis domain, with studies often lacking clarity around the data handling process.

Conclusions: A promising preliminary agreement between CRF predictions and measured values was noted. However, no definite conclusions can be drawn for clinical implementation due to high heterogeneity among the included studies and lack of external validation. Nonetheless, continuous data streams appear to be a valuable resource that could lead to a step change in how we measure and monitor CRF.

Trial Registration: PROSPERO CRD42024593878; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024593878

(JMIR Mhealth Uhealth 2026;14:e69996) doi:10.2196/69996

KEYWORDS

wearables; cardiorespiratory fitness; free-living data; machine learning; perioperative medicine

Introduction

Cardiorespiratory fitness (CRF) is regarded as a key element of anesthetic preassessment and preoperative decision-making, reflecting an individual's aerobic capacity and ability to withstand and recover from surgery. The most widely recognized measure of CRF is maximal oxygen uptake (VO_2max), a strong indicator of the cardiorespiratory system's ability to capture, transport, and use oxygen during exercise. VO_2max is inversely related to all-cause mortality and linked to several other health outcomes, such as cardiovascular disease (CVD), dementia, and depression [1-3]. In surgery, VO_2max serves as a prognostic marker and is currently the gold standard predictor of early postoperative cardiorespiratory morbidity [4,5].

Cardiopulmonary exercise testing (CPET) is a maximal dynamic test used to assess CRF and global exercise response. CPET is typically performed on a cycle ergometer or a treadmill under conditions of graduated physiological stress, involving computerized gas-exchange analysis of breath-by-breath ventilation. The test is routinely used in cardiorespiratory medicine as a diagnostic tool to distinguish between ventilatory and cardiac exercise intolerance [6]. In the preoperative setting, it is usually used as a risk assessment tool for major surgery to aid clinical decision-making regarding suitability for surgery and to guide perioperative management [7,8].

Despite its proven ability for risk stratification, there remain some drawbacks to this method. VO_2max measurements through maximal exercise can be strenuous and challenging for older or frail adults and those with musculoskeletal conditions who may be limited by pain rather than exertion [9]. Performance-reducing factors, such as peripheral arterial disease, osteoarthritis, and poor effort, have also been associated with inaccurate measurements, which may impact clinical decision-making [10]. In addition, high costs, the requirement for highly trained staff to undertake the test, and reduced hospital availability render regular VO_2max monitoring impractical.

To overcome these limitations, several VO_2max prediction models have been developed: nonexercise models are usually derived from lifestyle and anthropometric data, while submaximal exercise tests rely on prespecified protocols that involve heart rate (HR) monitoring at certain speeds, such as the 20-meter shuttle test or the modified shuttle walking test [9,11,12]. These methods offer an alternative; yet, they are not widely used in routine clinical practice due to some inherent limitations. Submaximal tests rely on the assumption that mechanical efficiency is the same for everyone, often leading to inaccurate VO_2 max estimations, and self-reported physical activity measures are subject to social desirability and recall bias [13]. Equally, lack of protocol standardization has raised concerns about the validity and reliability of submaximal tests [14]. Nonetheless, it seems a great limitation of the above CPET alternatives is their inability to capture and assess unstructured and incidental ambulatory activity accurately [15].

This gap has encouraged the exploration of wearable devices that can collect a substantial amount of information about an individual's activities in daily life, regardless of frequency, duration, or intensity [16]. Wearable technology has experienced a remarkable uptake over the last years, with more users appreciating the potential benefits for health and fitness tracking [17]. Commercially available wearables already offer VO_2max estimations; however, their algorithms are primarily based on short periods of structured exercise data, and their resulting VO_2max estimations have a large degree of error at the individual level [18]. Continuous monitoring of unstructured physical activity, however, shows promise in a variety of settings, enabling constant tracking of physiological data in an unobtrusive manner [16]. Physiological signals captured over longer periods may be more representative of CRF for certain populations. In view of the great potential of these devices, we aimed to explore whether CRF can be accurately predicted leveraging wearable data from unsupervised free-living conditions, outside the controlled laboratory environment. In this paper, we systematically review the research methodology behind the proposed models, the associated challenges and limitations, and discuss the feasibility of applying this concept for CRF estimation in health care settings.

Methods

Search Strategy and Study Selection Process

This study was registered with the international database for systematic reviews PROSPERO (CRD42024593878). We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement and recommendations for systematic reviews [19]. Relevant studies were located from a systematic electronic search of 4 databases (MEDLINE, Embase, Scopus, and arXiv), and the last search was performed on July 27, 2024. The full search strategy and key terms are available in Multimedia Appendix 1.

We used online systematic review software to blind reviewers and screen titles and abstracts after removing duplicates [20]. Conflicts were resolved through direct discussion between the 2 reviewers (AD and ABS) after unblinding. The full papers of potentially eligible studies were scrutinized against the eligibility criteria. Citation chaining of references was also completed by AD. Any additional studies identified were subsequently reviewed and assessed for inclusion by a third investigator (MRK; Textbox 1).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria:

- All papers reporting the development of a prediction model to estimate maximal oxygen uptake from longitudinal free-living wearable data.
- "Free-living" is defined as data collected in unsupervised, uncontrolled, real-world settings.
- Mixed designs with some simulated activities permitted, provided that at least some unsupervised activity was captured and analyzed.
- Human studies published in English.
- All wearable devices are eligible, including accelerometers, electrocardiogram (ECG) biosensors, commercial smartwatches, and optical heart rate sensors (photoplethysmogram).
- No restriction applied to the clinical setting studied.

Exclusion criteria:

- Studies in which the authors focused solely on physical activity and energy expenditure estimation.
- Monitoring occurring exclusively under controlled laboratory conditions, with no free-living activity studied.
- Wearable data including only exercise activity.
- Studies in which the authors did not report a prediction model but only correlations of wearable metrics with measures of cardiorespiratory fitness.
- Systematic reviews, literature reviews, surveys, conference proceedings, or meeting proceedings.
- Studies focusing on adolescents and young children.

We considered all papers reporting the development of a prediction model to estimate VO_2max from longitudinal free-living wearable data. For this review, "free-living" was defined as data collected in unsupervised, uncontrolled, real-world settings. Mixed designs with some simulated activities were also permitted, provided that at least some unsupervised activity was being captured and analyzed by the authors. A limit was set to human studies published in English. All wearable devices were eligible, including accelerometers, electrocardiogram (ECG) biosensors, commercial smartwatches, and optical HR sensors (photoplethysmogram). No restriction was applied to the clinical setting studied.

Studies were disqualified if (1) the authors focused solely on physical activity and energy expenditure estimation; (2) monitoring occurred exclusively under controlled conditions in a laboratory setting, and no free-living activity was studied; (3) wearable data included only exercise activity; (4) the authors did not report a prediction model, but only correlations of various wearable metrics with measures of CRF; (5) they were systematic or literature reviews, surveys, conference, or meeting proceedings; and (6) studies focusing on adolescents and young children.

Data Extraction and Model Performance Assessment

To ensure consistency, a standardized form was piloted and modified until consensus was reached between 2 authors and the senior investigator for the data extraction tool. Two reviewers (AD and ABS) retrieved all relevant data independently, and a third author (MRK) verified the accuracy of the records, cross-referencing with sources to resolve discrepancies. For each study, the following items were extracted: study details, demographics, setting and sample size, wearable device used for monitoring, and the baseline method used to obtain the ground truth (control). We recorded features derived from wearable data and the preprocessing techniques researchers used for feature extraction. We extracted details on the various machine learning (ML) models that were used, as well as prediction accuracy metrics and the validation process reported.

Quality Assessment

Qualitative appraisal of each included study was independently performed by 2 authors using a modified version of the Prediction model Risk Of Bias Assessment Tool (PROBAST) following the updated TRIPOD-AI (Transparent Reporting of a multivariable or ML prediction model for Individual Prognosis Or Diagnosis–artificial intelligence) guidance [21,22]. In case of disagreements, the opinion of the third author was sought.

Studies received a score of "low," "unclear," or "high" risk of bias on five major domains: (1) predictor choice and definition; (2) participant selection, including source and study setting; (3) outcome measurement; and (4) analysis and methodological quality of the proposed model. Overall judgment was rated as unclear if at least 1 domain was regarded as unclear, and similarly as high if any domain was rated as high. Risk-of-bias plots were created for quality assessment using the "robvis" software package in R (R Foundation for Statistical Computing).

Data Analysis

We calculated and reported descriptive statistics to outline each study characteristic. Summary measures are reported as means or medians, including measures of dispersion such as SDs. Key metrics of model accuracy were identified and combined for quantitative analysis.

Frequently reported metrics included the Pearson correlation coefficient (r) and the R^2 values. Other metrics such as the standard error of estimate (SEE) and the root-mean-square error were also reported but less frequently. Although not technically an accuracy metric, where available, the r coefficient was used to indicate how well the model's predictions aligned with the CPET values.

A meta-analysis of correlation estimates was undertaken to integrate measures of performance across the included studies and provide a more objective and systematic assessment. In this instance, reported correlation coefficients (r) were used as the primary effect size, as they represented the most consistently reported metric in this review. When only R^2 values were available, we converted them by taking the square root to obtain the corresponding r values between the predicted and actual VO_2max values and assessing the direction of association in each study. RStudio (version 2024.04.2+764; R Foundation for Statistical Computing) was used for all statistical analyses [23]. Two packages, "metafor" and "robumeta," were installed to perform the meta-analysis. Fisher Z transformation was applied to convert correlation estimates to a more normally distributed metric and obtain standard effect sizes. A restricted maximum likelihood estimation method was used for a standard random-effects model to conduct the meta-analysis [24]. The random-effects model assigns less study weight to larger studies with less variance [24]. Results of the meta-correlation analysis were presented visually using a forest plot. Subgroup analysis was also performed, comparing regression-based models with more advanced ML methodologies.

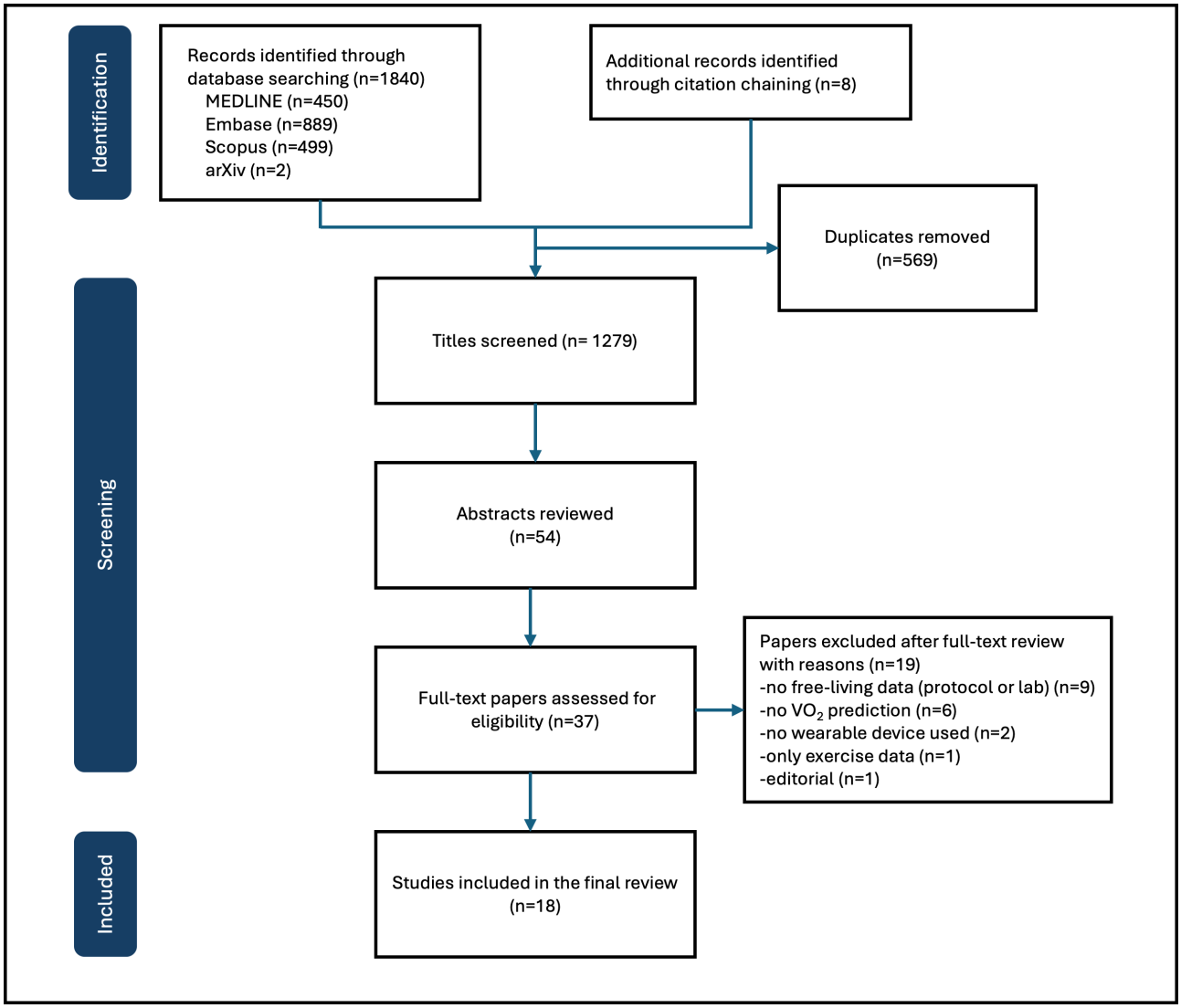
We assessed heterogeneity using the I^2 and τ^2 statistics. The I^2 statistic quantifies the proportion of total variation in effect sizes that was due to heterogeneity rather than chance. We considered I^2 values of 25%, 50%, and 75% to represent low, moderate, and high heterogeneity, respectively [25]. The τ^2 represents another method to assess the between-study variance, focusing on the absolute variability of true effect size, with higher values indicating greater heterogeneity [26]. The Egger test was used to assess the likelihood of publication bias, which was also presented visually with a funnel plot.

Results

Overview

The combined literature search generated 1848 papers, with 1279 remaining after deduplication. The PRISMA flowchart (Figure 1) shows the paper selection process. Following title and abstract screening, 37 papers qualified for full-text review. Eighteen studies were accepted in the final set with a total sample size of 31,072 participants. Sample sizes varied greatly across studies and ranged from 13 to 12,425. Only 4 studies had a sample size of more than 1000 patients, indicating that most research in this field is based on a relatively small number of participants [27-30].

Figure 1. Flowchart of study selection following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidance. VO_2: maximal oxygen uptake.



The characteristics of each included study are shown in Table 1 [2,4,5,7,11,12,16,27-37]. Participant-level weighted mean age was 46.9 (SD 1.46) years, with a male participant distribution of 48.9%. Notably, a few studies had exclusively male or female participants, while others had more balanced samples. Most studies included data from volunteers recruited in a prospective manner (n=13, 72%). Smaller studies focused primarily on healthy participants (n=13), in contrast to the larger cohorts (n=5) that used data from population studies involving hundreds of patients (the Fenland and Framingham studies) [38,39]. Overall, there were only 2 research trials that targeted patients scheduled for preoperative assessment [4,7].

Table . Summary of study characteristics.

Authors	Year	Sample size (M[a]), n (%)	Age (years), mean (range)	Sensors and modality used	Wearable monitoring (days)	Participants	Control	Reference VO_2, mean (SD)
Plasqui and Westerterp [2]	2005	25 (40)	28 (18-50)	Tracmor elastic belt triaxial accelerometer and Polar (S610i) HR[b] monitor wristwatch	7 (daytime only)	Healthy volunteers	Maximal GXT[c] cycle	M=49.5±10.2; F[d]=40.7±8.4
Plasqui and Westerterp [33]	2006	26 (53.8)	29 (18-50)	Tracmor elastic belt triaxial accelerometer and Polar (S610i) HR monitor wristwatch	7 (daytime only)	Healthy volunteers	Maximal GXT cycle	44.6±10.5
Cao [34]	2009	189 (0)	49.6 (20-69)	Kenz Lifecorder uniaxial accelerometer	7 (daytime only)	Healthy volunteers	Maximal GXT cycle	31.4±7.4
Cao et al [35]	2010	148 (0)	47 (20-69)	Kenz Lifecorder uniaxial accelerometer and triaxial accelerometer	7 (daytime only)	Healthy volunteers	Maximal GXT cycle	30.8±5.9
Novoa et al [4]	2011	38 (79)	62.8 (38-80)	OMROM Walking Style ProW accelerometer (two modes-aerobic mode after 10 min of walking at 60 steps per min)	7-41 (daytime only)	Patients scheduled for lung resection	Maximal GXT cycle and Arterial gas	20.3±4.6
Altini et al [11]	2016	46 (45)	24.7 (NR[e])	ECG[h] Necklace (one lead ECG) and ADXL330 triaxial accelerometer and Mobile phone for GPS coordinates	14 (during laboratory protocols and free-living)	NR	Maximal GXT cycle	44±9.8
Altini et al [5]	2016	51 (47)	NR[e]	ECG Necklace (one lead ECG) and ADXL330 triaxial accelerometer	14 (during laboratory protocols and free-living)	Healthy volunteers	Maximal GXT cycle	NR[e]
Beltrame et al [37]	2017	13 (100)	26 (NR)	Hexoskin smartshirt (hip accelerometer, three lead ECG and respiration bands)	4 (free-living 9 AM to 5 PM) and simulated ADLs[f]	Healthy volunteers	Simulated ADLs and Pseudorandom Ternary sequence.	NR

Authors	Year	Sample size (M[a]), n (%)	Age (years), mean (range)	Sensors and modality used	Wearable monitoring (days)	Participants	Control	Reference VO_2, mean (SD)
Ahn et al [31]	2017	24 (100)	27.5 (NR)	Shimmer ECG sensor (18), for measurement of 2-lead ECG and triaxial accelerometer	4 (daytime only)	Healthy volunteers	Maximal GXT treadmill	48.5±5.3
Kwon et al [12]	2019	240 (47)	42 (20-65)	Fitbit (Fitbit Charge; Fitbit) (triaxial accelerometer, PPG)[g]	3 (daytime only)	Healthy volunteers	Maximal GXT treadmill	36.25
Bonomi et al [16]	2020	40 (48)	25 (18-55)	Chest-belt HR ECG monitor (RS800CX, Polar, Wrist activity monitor (Tracmor)	5 (daytime) and simulated ADLs	Healthy volunteers	Maximal GXT cycle	M=45.7±6.1; F=40±6.6
Jones et al [7]	2021	49 (65)	65 (NR)	Garmin Vivosmart HR+ activity tracker wristwatch (PPG, accelerometer, GPS)	7 continuous	Patients for preoperative assessment	Maximal GXT cycle	18.2±4.5
Spathis et al [30]	2021	2100 (46)	48.7 (35-65)	Combined HR and uniaxial movement sensor (Actiheart) and wrist triaxial accelerometer	6 continuous	Population-based cohort study (Fenland)	Submaximal GXT treadmill	NR
Wu et al [28]	2022	12,425 (47) 181	47.7 (35-65)	Combined HR and uniaxial movement sensor (Actiheart) and wrist triaxial accelerometer	6 continuous	Fenland study population-based cohort and UK Biobank	Submaximal GXT treadmill Maximal GXT test	NR; 32.95
Spathis et al [27]	2022	11,059 (47)	47.7 (35-65)	Combined HR and uniaxial movement sensor (Actiheart) and wrist triaxial accelerometer	6 continuous	Fenland study population-based cohort	Submaximal GXT treadmill	M=41.95±4.61; F=37.44±4.73
Frade et al [32]	2022	43 (74.4)	37.5 (19-72)	Hexoskin smartshirt (hip accelerometer, three lead ECG and respiration bands)	7 (daytime only)	Volunteers (chronic disease allowed)	Maximal GXT cycle	32.09
Neshitov et al [29]	2023	3894 (67)	42 (20-65)	Apple Watch and Garmin watch (PPG, triaxial accelerometer, GPS)	mean 287, SD 149	Healthy volunteers -consented to Welltory app	Estimated by smartwatch device	36.16±6.66

Authors	Year	Sample size (M[a]), n (%)	Age (years), mean (range)	Sensors and modality used	Wearable monitoring (days)	Participants	Control	Reference VO_2, mean (SD)
Zhang et al [36]	2024	662 (41)	53 (NR)	Apple Watch (PPG, triaxial accelerometer, GPS)	mean 128 (daytime only)	Framingham study cohort	Maximal GXT cycle	M 27±7; F 22±6

[a]M: male.

[b]HR: heart rate.

[c]GXT: graded exercise test.

[d]F: female.

[e]NR: not reported.

[f]ADL: activities of daily living.

[g]PPG: photoplethysmogram.

[h]ECG: electrocardiogram.

A wide variety of wearable devices were used, including triaxial accelerometers, ECG sensors, and smartwatches with optical photoplethysmogram sensors. All studies used accelerometers for motion tracking, while HR monitoring was performed either with ECG (n=9) or optical sensors (n=6). The monitoring durations also differed, with most studies tracking participants over a few days, usually 3 - 7, while others extended up to nearly a year [29]. Measurement of VO_2max as the ground truth was typically obtained through a maximal CPET test, though in 4 trials, a submaximal treadmill test was used. In 1 study, reference VO_2max was not directly measured but estimated using the proprietary algorithm of a smartwatch device [29]. Due to disparities in participant populations, the reference mean VO_2max varied significantly between patient-centered studies (weighted mean VO_2max 19.2 mL/kg/min, SD 1.48) and healthy volunteer studies (40.2 mL/kg/min, SD 6.58; z test, P=.03).

Features

Feature extraction is a crucial first step in signal processing, transforming complex raw data points into meaningful numerical features that can be interpreted and processed in a model [40]. In high-volume continuous wearable data, feature extraction can also help to reduce dimensionality without losing important information. This process makes data handling easier and speeds computation by focusing only on the most relevant aspects of the data [41]. We observed, however, that in 6 studies, researchers adopted features as reported from the internal proprietary algorithm of the manufacturer without further analysis of the raw data. Table 2 summarizes the analytical methods used in the included studies, while Table 3 provides an overview of each feature type with examples in the studies used.

Table . Summary of analytical methods, models, and wearable features used in the included studies.

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Plasqui and Westerterp [2], 2005	[a]HR/ACM[b] index (ACM: activity counts per minute)	Multiple linear regression	Minute averages for HR and activity counts averaged over the 7 days, ratio of HR/ACM, and missing data removed	SEE[c]=12.4%; r=0.87	NR[d]	Fitness index HR/ACM significantly related to VO_2max[e] corrected for body composition and age
Plasqui and Westerterp [33], 2006	HR/ACM index	Multiple linear regression 2nd equation tested	Same as above and groups combined and sorted for activity counts	r=0.86; SEE=10.7%; Bland-Altman systematic error=5.6%	Cross-validation	A second equation for the fitness index HR/ACM had to be tested to predict VO_2[f]
Cao et al [34], 2009	Daily SC[g]	Hierarchical linear regression	Steps per day provided and handling of missing data not reported	SEE=10.9%; r=0.81	Split test	SC was a significant contributor to the prediction of the measured VO_2max
Cao et al [35], 2010	MVPA[h], VPA[i], and SC	Hierarchical linear regression	As provided (minutes spent in MVPA and VPA) and handling of missing data not reported.	SEE=9.66%; r=0.863	Cross-validation, Subgroup analysis	VPA significantly increased the explained variance in VO_2max, adjusted for age
Novoa et al [4], 2011	Daily SC, Aerobic SC, time spent in aerobic activity, and daily distance in km	Linear regression	As provided and handling of missing data not reported	R^2=0.93	Bootstrapping (1000 iterations)	VO_2max can be significantly predicted by the mean daily walked distance
Altini et al [11], 2016	HR at different walking speeds, stay regions, and activity composites (ie, HR at relative time spent in each activity)	Nonnested hierarchical Bayesian regression, SVM[j] classifier, and HMM[k] for transitions between activities	LDA[l] activity is classified into 6 clusters, accelerometer data was bandpassed between 0.1 and 10 Hz to isolate dynamic components, HR was extracted from R-R intervals and averaged over 15 seconds, and missing data not analyzed	R^2=0.76; RMSE[m]=249.4; SEE=5.79%	Leave-one-participant-out cross-validation	Contextualizing HR by means of activity and speed improved correlation between free-living HR and CRF[n]
Altini et al [5], 1985	HR/min while lying down and while walking at 3.5 and 5.5 km/h	Multiple Linear regression, SVM classifier	HR was extracted from R-R intervals and averaged over 15 seconds windows and unusable ECG[o] was discarded. Acceleration signal was segmented in nonoverlapping intervals of 5s	R^2=0.78; RMSE=284.7	Leave-one-participant-out cross-validation	Submaximal context-specific HR can be used to estimate VO_2max

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Beltrame et al [37], 2018	Means of HR, VE[p], BF[q], hip acceleration, and SC in the 2 conditions ("active-inactive")	Random forest	HR was averaged every 16 beats, VE and BF average of the last 7 respiration cycles, all features were time-aligned, low-pass filtered at 0.01 Hz. Fast Fourier transformation and frequency domain analysis for hip acceleration, and when hip acceleration was >0.05 g, data were labeled as "active"; otherwise, they were "inactive	r=0.88	Leave-one-participant-out cross-validation	Predicted oxygen uptake data during ADLs[r] were strongly correlated with the temporal characteristics of the VO_2 during a controlled protocol
Ahn et al [31], 2017	aEE[s] (nonlinear model derived from ACM horizontal and ACM vertical signals and HR per minute	Linear regression between HR and aEE	The tri-axial acceleration was band-pass filtered (0.25 to 7 Hz). ECG data were band-pass filtered (5 to 20 Hz), the R-R intervals were averaged for 1 min and converted to a HR, and used only increasing HR periods and excluded data with inaccurate ECG	R^2=0.74; r=0.87; SEE=11.85%	Split test	aEE can be used to estimate VO_2max during daily activities
Kwon et al [12], 2019	HR and daily PA in terms of METs, used slope of HR and PA	Linear regression	Moving average filter was applied, data points at which both HR and physical activity data increased were selected	R^2=0.651; SEE=3.518; 9.6%	PRESS[t] for cross-validation	VO_2max can be estimated using novel features, aEE and the slope between physical activity and HR
Bonomi et al [16], 2020	Acceleration and HR- fitness index named TEE-pulse[u]	Stepwise linear regression	Motion intensity was defined as activity counts per minute, acceleration signal was processed in overlapping windows of 60 seconds, and activity is grouped in (sedentary, or other) based on a set of counts thresholds	RMSE=367 or 12.4%; SEE=13.09%; r=0.89; MAE=10.2%	Leave-one-participant-out cross-validation	The daily average TEE-pulse was highly correlated to the mean TEE-pulse measured in the laboratory without the need for specific exercise protocol
Jones et al [7], 2021	Floors climbed, total number of steps and total distance, average HR and resting HR	Linear regression	Features were used as provided by the device and averaged across the 7-day wear period. Self-reported METs[v] from questionnaires	AIC[w]=181.62; R^2=0.74; r=0.86; AUC[x]=0.93	NR	Using all the wearable variables together in linear regression gave a stronger correlation between the measured CPET[y] values, specifically for peak VO_2

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Spathis et al [30], 2021	HR per minute, acceleration (magnitude calculated through ENMO[z]), resting HR	Step2Heart Deep neural network, (CNN[aa] learn spatial and RNN[ab] temporal features)	Noisy heart data removed with a Gaussian process robust regression, participants with less than 72 hours of wear were removed, and accelerometry and ECG signals were summarized to a common time resolution of one observation per 15 seconds	AUC=0.70; RMSE=9.54 (HR forecasting only)	Split test	A general-purpose self-supervised feature extractor for wearable data was developed. HR forecasting transfer learning of learned physiological representations
Wu et al [28], 2022	HR and movement 26 features combining HR, movement data, and time-series metadata	UDAMA[ac]	Nonwear periods were removed (periods of nonphysical HR and no movement), downsampled the sampling rate to 15 minutes and used the first 600 timesteps, and pretrained on noisy data and used adversarial training on the BBVS dataset	R^2=0.392; r=0.665; MSE[ad]=30.79; MAE[ae]=4.44	Split test and 3-fold cross validation	A novel model proposal to leverage noisy data from source domain (wearable dataset) to improve modeling for accurate fitness estimation at scale
Spathis et al [27], 2022	48 features: Raw acceleration derived through ENMO, HR, HRV[af], MVPA for each feature mean, minimum, maximum, SD, and the slope of a linear regression fit	Deep neural network- adaptive representation learning	Nonwear periods removed, movement intensities were converted into standard METs, principal component analysis for noise reduction, tSNE[ag], a nonlinear dimension-reduction technique was applied	R^2=0.658; RMSE=2.956; r=0.82; RMSE=8.998 (Biobank only)	Split test and External validation in UK biobank cohort (181 patients)- Maximal GXT[ah] testing	A deep learning framework for predicting CRF was developed, combining learned features from HR and accelerometer free-living data without context awareness
Frade et al [32], 2023	Mean HR and BF, minute ventilation, tidal volume, mean hip acceleration, and mean SC	SVM (support vector regression formulation)	Abnormal HR and BR[ai] were excluded with a preprocessing algorithm (not mentioned), and all raw data were averaged	r=0.804; MAE=3.84	k-fold cross-validation	Hemodynamic domain presented statistically higher importance to predict the VO_2max compared with activity and Pulmonary domains

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Neshitov et al [29], 2023	HR and SC/min, HR/cadence ratio, daily MET, and HR response to cadence increase	Quantile regression (for each quantile a gradient boosting model was trained)	The HR stream was resampled to 1 measurement per minute and averaged over consecutive 1-minute intervals, cadence is the number of steps made during the same 1-minute interval, continuous ranked probability score used for hyperparameter tuning, and model trained on estimated VO_2 from wearable device	Test set: ECE[aj]=0.032; IQR=3.948; MedPE[ak]=0.01; and Direct VO_2 dataset: ECE=0.084; IQR=4.705; MedPE=0.35	Split test for training External validation (10 healthy volunteers) Maximal GXT treadmill	Anthropometric characteristics were the most influential feature, followed by cadence to HR ratio. The proposed model provides a point estimation and a probabilistic prediction of VO_2; to estimate the prediction's uncertainty

Study, year	Wearable features	Model used	Preprocessing techniques and analysis	Model performance	Validation	Principal finding
Zhang et al [36], 2024	Daily SC, mean HR	Multivariable linear regression, sensitivity analysis for age, BMI, gender	Defined HR as nonactive if recording interval was >1 minute, hourly steps <30, inferred motion context status for HR measures that lacked motion, and excluded days with <5 hours of wearing time	R^2=0.07 - 0.12	NR	Every 1.3 mL/kg/min higher peak VO_2 corresponded to a 2.4-bpm lower nonactive HR. Physical activity with 1.3 mL/kg/min higher peak VO_2 was associated with nearly 1000 more daily steps

[a]HR: heart rate.
[b]ACM: activity counts per minute.
[c]SEE: standard error of estimate.
[d]NR: not reported.
[e]VO_2max: maximal oxygen uptake.
[f]VO_2: measured oxygen uptake.
[g]SC: step count.
[h]MVPA: moderate to vigorous physical activity.
[i]VPA: vigorous physical activity.
[j]SVM: support vector machine.
[k]HMM: hidden Markov model.
[l]LDA: latent Dirichlet allocation.
[m]RMSE: root-mean-square error.
[n]CRF: cardiorespiratory fitness.
[o]ECG: electrocardiogram.
[p]VE: minute ventilation.
[q]BF: breathing frequency.
[r]ADL: activities of daily living.
[s]aEE: activity energy expenditure.
[t]PRESS: predicted residual error sum of squares.
[u]TEE: total energy expenditure.
[v]MET: metabolic equivalent task.
[w]AIC: Akaike Information Criteria.
[x]AUC: area under the receiver operating characteristic curve.
[y]CPET: cardiopulmonary exercise testing.
[z]ENMO: Euclidean norm minus one.
[aa]CNN: convolutional neural network.
[ab]RNN: recurrent neural network.
[ac]UDAMA: unsupervised domain adaptation via multidiscriminator adversarial training framework.
[ad]MSE: mean squared error.
[ae]MAE: mean absolute error.
[af]HRV: heart rate variability.
[ag]tSNE: t-distributed stochastic neighbor embedding.
[ah]GXT: graded exercise test.
[ai]BR: breathing rate.
[aj]ECE: expected calibration error.
[ak]MedPE: median prediction error.

Table . Wearable features grouped by type, with examples used in each of the included studies.

Feature type	Studies	Examples
Motion	• Cao et al [34] and Cao et al [35] • Novoa et al [4] • Beltrame et al [37] • Jones et al [7] • Spathis et al [27] and Spathis et al [30] • Frade et al [32] • Zhang et al [36]	• Daily step count • Moderate-to-vigorous physical activity • Vigorous physical activity daily distance, mean hip acceleration, and acceleration magnitude
Cardiac	• Altini et al [11] • Beltrame et al [37] • Ahn et al [31] • Jones et al [7] • Spathis et al [30] • Frade et al [32] • Zhang et al [36]	• Average HR[a] and resting HR • HR/min • Heart rate variability measures
Contextualized HR	• Plasqui and Westerterp [2] and Plasqui and Westerterp [33] • Altini et al [5] and Altini et al [11] • Bonomi et al [16] • Kwon et al [12] • Wu et al [28] • Spathis et al [27]	• HR/ACM[b] ratio index • HR response to cadence increase activity composites (HR at relative time spent in an activity) • HR at different walking speeds • slope of HR and physical activity • HR/cadence ratio
Other	• Beltrame et al [37] • Ahn et al [31] • Kwon et al [12] • Frade et al [32] • Neshitov et al [29]	• V_E[c], BF[d], and tidal volume • aEE[e] • TEE[f] • MET[g]

[a]HR: heart rate.

[b]ACM: activity counts per minute.

[c]VE: minute ventilation.

[d]BF: breathing frequency.

[e]aEE: activity energy expenditure.

[f]TEE: total energy expenditure.

[g]MET: metabolic equivalent.

Motion Features

Activity features from the accelerometer data included mainly daily step count (SC), distance covered, time spent in anaerobic or sedentary activity (stay regions), and acceleration or walking speed (Table 3). From these, the most frequently reported feature was the daily SC, which in most instances was precalculated from the accelerometer's proprietary algorithm. We found that some researchers reported steps as an average across several days, removing the temporal context which might be useful in analyzing trends [2,4]. Intensity of movement and walking speed was described as time spent in sedentary or vigorous activity, which can be calculated from the acceleration as activity counts per minute [16]. One study used simulated daily activities to correlate aerobic dynamics of variable intensity [37]. Finally, distance covered was either extracted directly from the device or computed as the total number of steps taken in a day multiplied by the stride length of the participant [4].

Cardiac Features

Features extracted from the ECG and optical sensors included average HR, resting HR, HR per minute, mean HR, and ΔHR (difference between current and previous HR value to capture the magnitude of changes in cardiac activity). ECG signals are generally complex; they are subject to motion artifact, creating noise that affects quality [42]. Various filtering methods were applied to reduce noise. Researchers commonly use a band-pass filter between 5 and 10 Hz to remove artifacts and enhance the detection of the heartbeat from the R-R intervals. The R-R intervals were usually averaged over set-time windows, discarding any inaccurate values [11,31]. Beltrame et al [37] averaged HR every 16 beats, passing HR data through a low-pass filter at 0.01 Hz, removing high frequencies.

Contextualized HR

The combination of HR and activity data, often referred to as contextualized HR, was reported in more recent studies. This contextual dimension of wearable signals can be vital in understanding the physiological cardiac response to exercise. Such features comprised HR at variable-intensity walking speeds, HR and SC per minute, HR/cadence ratio, and HR response to cadence increase (Table 2) [5,27,29]. Kwon et al [12] examined the slope between the concurrent increase in physical activity and HR, while others studied time-series

metadata of HR and movement signals, mining numerous features [27,28]. Advanced signal processing methods, such as principal component analysis and fast Fourier transform, were implemented to tackle noisy heart data, but this was not standardized across the included studies. Authors frequently used resampling techniques (standardizing time intervals between data points) to align HR with movement data, helping to contextualize HR within the corresponding physical activity [30].

Other Features

Less common features included energy expenditure estimates in terms of metabolic equivalents, which were calculated in 3 studies based on daily physical activity and proprietary algorithms [12,16,29]. On 1 occasion, breathing frequency and minute ventilation were extracted as the average of the last 7 respiration cycles, based on respiration bands integrated into the wearable device used for monitoring [37].

Models

Multiple modeling techniques were used among the included studies, with more advanced ML models such as support vector machines (SVMs) and deep learning gaining interest over regression in recent studies. This trend follows the usage of preprocessing techniques such as fast Fourier transform and frequency domain analysis and represents an effort to mine the raw data and uncover hidden patterns. A detailed breakdown of modeling and preprocessing techniques is provided in Table 2.

Eleven studies used linear models, which allow interpretation of outcomes through reporting of coefficients that give direct insights into how each predictor influences the outcome measured [2,4,7,11,12,16,31,33-36]. The earliest study that examined whether the ratio of HR to activity counts could predict VO_2max was from Plasqui and Westerterp [2]. Two studies used correlation analysis to identify the strongest predictors of VO_2max and built on this using linear regression [4,7]. The rest of the studies outlined combinations of modern ML techniques. Three studies reported SVMs, with Frade et al [32] presenting support vector regression, which is an SVM formulation for regression problems. SVMs are supervised models that identify an optimal decision boundary (hyperplane) to separate data points into distinct classes with the aim of maximizing the margin between observed and predicted values.

Beltrame et al [37] were the only group to consider a random forest model. Random forests aggregate several decision trees together as a group but introduce randomness to prevent overfitting [37]. Another study trained several gradient boosting models and then fitted a quantile regression algorithm to predict the distribution of VO_2max with CIs [29]. In boosting, models are trained sequentially, building on the errors or residuals of the previous model to improve their prediction accuracy.

Finally, 3 studies [27,28,30] leveraged large-scale free-living datasets to predict CRF with variations of neural networks and deep learning. Wu et al [28] introduced a novel 2-stage approach building an adversarial training framework based on unsupervised domain adaptation. The proposed model was pretrained with noisy health-related labels in a fully supervised setting to improve its performance on high-quality, gold-standard data. Coarse- and fine-grained discriminators were used to better handle the distribution shifts between source (silver-standard) and target (gold-standard) datasets. Spathis et al [27] applied principal component analysis to denoise the raw data and developed deep neural network models able to capture nonlinear relationships between numerous wearable features and VO_2 max.

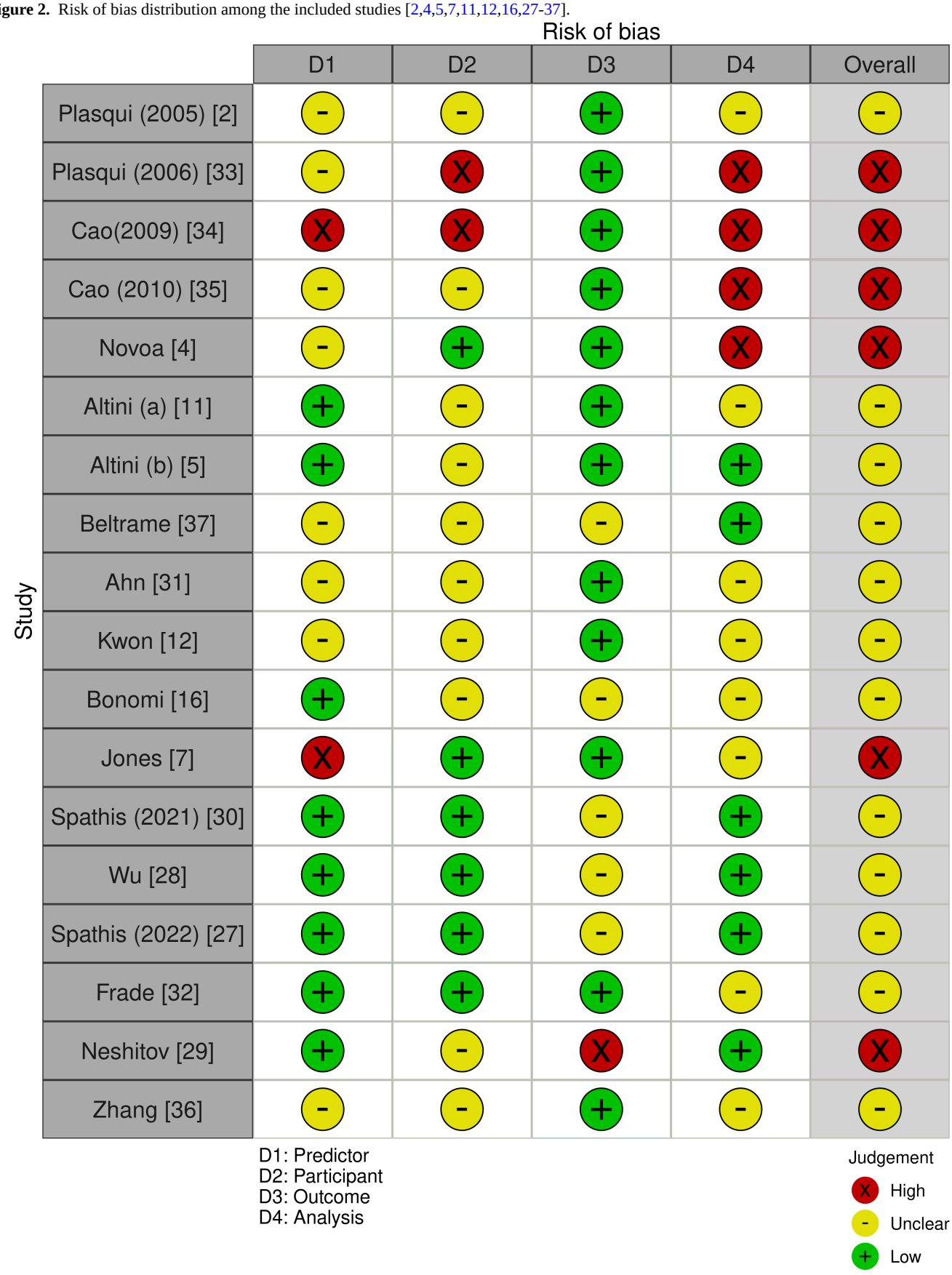
Length of Available Data Required for Prediction

Some studies examined the minimum length of free-living wearable data that would be required to reach reliable conclusions regarding VO_2max estimations, but no agreement was observed. The most thorough assessment was provided by Neshitov et al [29], who tested the degree of certainty for 5 different models using various amounts of available data. They advocated that a minimum of 200 minutes is required for an error estimation range of 4.5 mL/kg/min, but more than 1000 minutes is needed to improve this to under 4 mL/kg/min. Ahn et al [31] plotted the correlation coefficient values with the included measurement time and found no drastic improvements between VO_2 max and estimated values past the 900-minute mark, which yielded an r value of 0.81. On the contrary, other researchers reported that even 10 minutes per day of good-quality data might be sufficient to predict VO_2max. Beltrame et al [37] determined the 10-minute window from the frequency domain analysis as the ideal size for data extraction based on iterative testing of different window lengths (200-1000 seconds, incrementing by 100 seconds). A window length of 600 seconds (10 minutes) was found to provide the best balance between maximizing frequency resolution and ensuring enough reliable samples for analysis across participants. Altini [5,11] proposed this on a theoretical basis, as many submaximal protocols are of a 10-minute duration (eg, 6-minute walking test). Other studies did not account for a minimum length but excluded patients with <72 hours of data from the analysis [27]. Ultimately, this large disparity in the length of data required for feature engineering reflects the exploratory nature of some of the included studies.

Quality of Studies

Risk of bias distribution for each domain is provided in Figure 2 [2,4,5,7,11,12,16,27-37]. Critical appraisal of the included papers showed that only 1 study was classified as “low risk of bias” for all domains. CRF was aptly measured as the outcome of interest in 12 (67%) studies using data from maximal exercise tests (also provided in Multimedia Appendix 2). Appropriate selection of participants and predictors was documented in 6 and 8 studies, respectively. Higher degrees of bias were observed in the analysis domain with robust reporting of analytical methods noted only in 6 (33%) studies. Handling of missing data was not reported adequately in 5 studies [4,5,11,34,35], while others excluded data from analysis arbitrarily, without fully justifying their decisions [12].

Figure 2. Risk of bias distribution among the included studies [2,4,5,7,11,12,16,27-37].

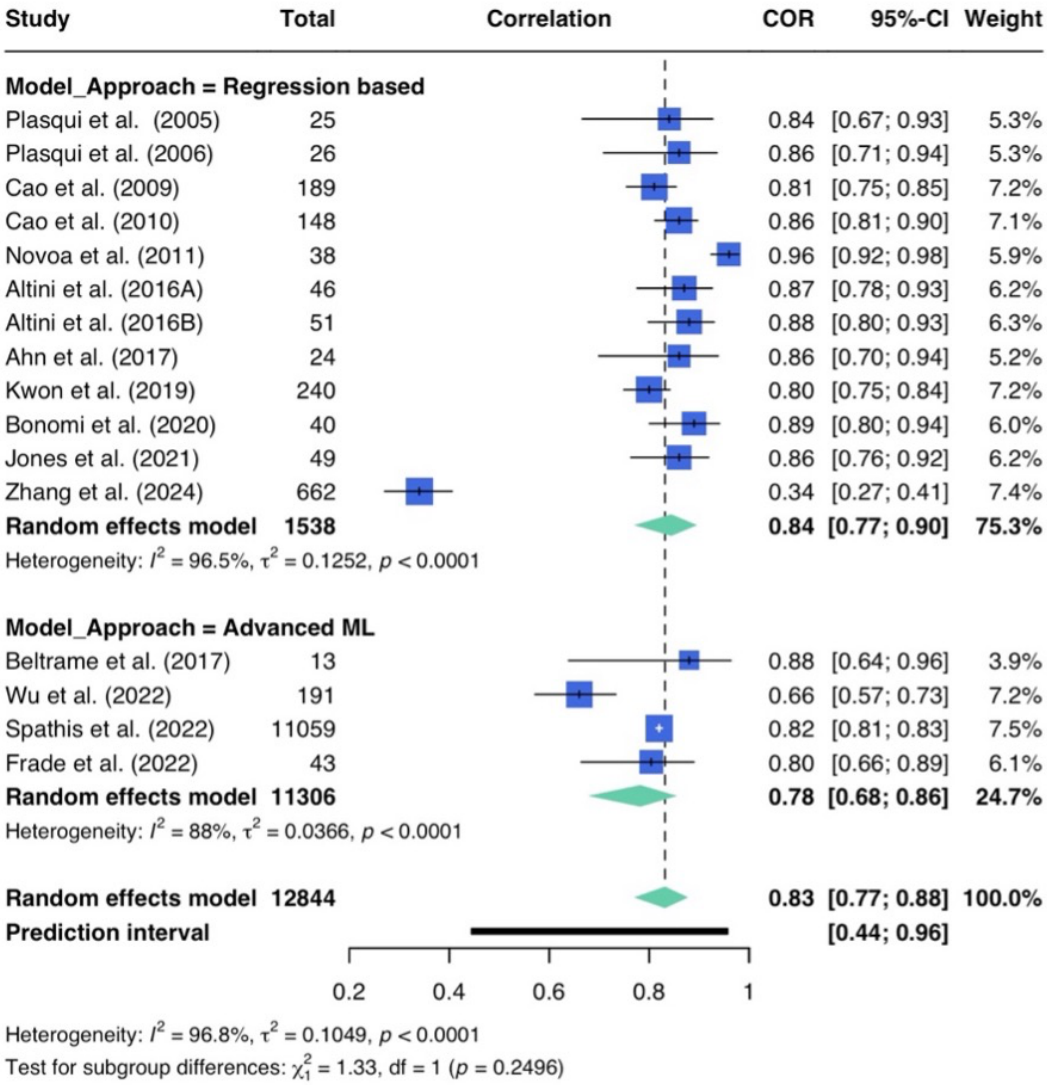


Most studies (n=13) reported internal validation methods for their predictive model, such as split-test or leave-one-participant-out cross-validation (Table 2). Model validation was not considered in 3 studies, and only 2 papers tested their algorithm externally on unseen data [27,28].

Model Performance

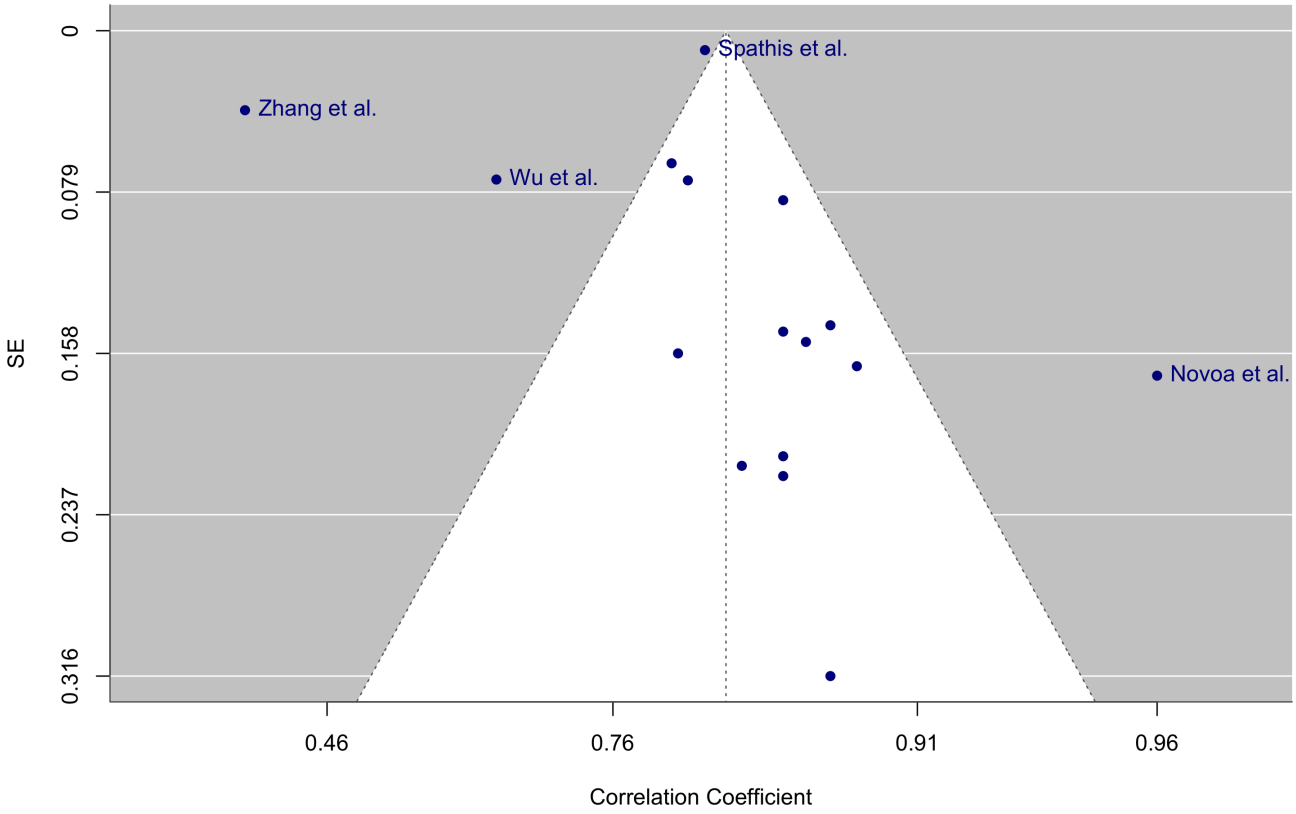
Various model performance metrics were reported (Table 2). The weighted average SEE was 9.03%, indicating that models overall predict VO_2max with an error of approximately 9%. In total, 16 studies were included in the meta-correlation analysis, which is provided in Figure 3 [2,4,5,7,11,12,16,27,28,31-37]. The pooled overall estimate of r=0.83 with a 95% CI of 0.77-0.88 from the random-effects model indicates a positive agreement between predicted and observed VO_2 max values. Heterogeneity among the included studies was high, with I^2=97% and a Q test of statistical significance (P<.01). Furthermore, moderate variance was observed (τ^2=0.1049), suggesting underlying differences in how well VO_2 max is predicted across studies. Subgroup analysis comparing regression-based methods with more advanced ML methodologies favored regression, but the difference was not statistically significant (P<.24; Figure 3 [2,4,5,7,11,12,16,27,28,31-37]).

Figure 3. Forest plot of the meta-correlation analysis between maximal oxygen uptake estimates and reference values. Random-effects study weighting was calculated as the inverse sum of the in-study variance and the between-study variance (τ^2). Subgroup analysis comparing modeling approaches is also presented [2,4,5,7,11,12,16,27,28,31-37].



In studies reporting high correlations, there are several common features. While most research indicated the use of validation methods with unseen data (train-test split and cross-validation) to test model performance, several studies reported the R^2 values from the linear regression model [4,7]. Another common factor among the highest-performing models was the incorporation of data collected from laboratory protocols, either to contextualize or interpret free-living data, into feature extraction and modeling processes [5,11,16]. In addition, the funnel plot provided in Figure 4 [4,27,28,36] revealed asymmetry with 4 outliers, which was also confirmed with an Egger test (z=2.29; P=.02). This suggests the presence of small-study effects and publication bias. Notably, the overoptimistic, smaller-sized study by Novoa et al [4] is likely overrepresented in the pooled estimate, and results need to be interpreted with caution.

Figure 4. A funnel plot illustrating the distribution of study effect sizes to assess potential publication bias. Four studies fall outside the expected range, suggesting potential publication bias [4,27,28,36].



Discussion

Principal Findings

This systematic review identified research using real-world, unsupervised wearable data to develop predictive models for CRF estimation, focusing on VO_2 max as the measure of interest. Our study adds to the literature as the first to appraise evidence in this field and showcase the ability of advanced ML algorithms to harness the power of unstructured physical activity outside controlled laboratory settings. The included meta-correlation analysis revealed a pooled overall estimate of 0.83 with a 95% CI of 0.77-0.88, and a mean SEE of 9.06%, demonstrating a promising overall agreement between predicted VO_2max and ground truth. Authors experimented with a range of sensor modalities and various population groups, but models were predominantly designed based on small-sized, healthy volunteer data. Several features were extracted from the free-living information, including SC and distance covered, resting and mean HR, and cardiac response to cadence increase. Quality control of the eligible studies showed that authors were consistent in predictor and outcome reporting, but analytical methods were often ambiguous and included some arbitrary decisions regarding data manipulation.

Advantages

The concept of CRF estimation based on free-living activity holds considerable potential, and results from this study suggest that this could be a pragmatic alternative to CPET. Leveraging longitudinal wearable data can aid preoperative risk assessment for frail patients or those with musculoskeletal conditions that underperform during CPET (indicated usually by a respiratory exchange ratio of <1.10) [8]. Researchers have argued that physiological signals captured over longer time periods may even be more representative of cardiac health in these populations [5,37] compared to a snapshot laboratory measurement. In addition, at-home monitoring offers a convenient and unintrusive assessment without the need for specific protocols, improving patient experience and reducing the psychological stress related to the hospital environment [43]. Considering decreasing costs and increasing accessibility [44], continuous monitoring could represent a complementary, more cost-effective, efficient method that can be scaled to accommodate all patients [7]. Serial measurements of VO_2max can not only help patients track progress and meet targets set during prehabilitation and rehabilitation programs but also guide clinician decision-making [45].

CRF is a well-established marker of CVD and all-cause mortality [3]. Considering that low levels of CRF may precede the clinical detection of CVD, early recognition and intervention are of patient benefit [37]. Wearable-driven evaluation of the aerobic response during unsupervised activities of daily living holds prognostic value in tracking changes in fitness over time, as demonstrated by Spathis et al [27]. Arguably, models that predict future CRF levels could help identify early-stage CVD before general symptom manifestation [37]. Finally, scrutinizing free-living data with advanced ML presents a rare opportunity to study patient behavior and activity habits, shedding light on individual CRF levels [37]. But above all, tailored interventions can be implemented promptly to improve patient fitness and outcomes [46].

Contextualizing HR

The advent of wearables is undoubtedly transforming the landscape of health monitoring, providing clinical teams with a substantial amount of user-generated time-series data [27]. Although some earlier published studies used aggregates over several days as features (average steps or HR data), potentially losing information on trends and variability [2,33-36], most researchers in this review worked on extracting features from the raw signals, with seven studies aligning cardiac and activity points to contextualize HR and gain insights into the participants' physiological response to workload. Based on this principle, Plasqui and Westerterp [2] were the first to publish a fitness index as a ratio between acceleration and HR. Studies of more advanced modeling included other multimodal features such as HR at certain speeds or activities and the HR response to acceleration and recovery [5,29,37]. As physical activity encompasses both body movement and an associated cardiovascular response, leveraging these signals concurrently allows for a better evaluation of the temporal dynamics of CRF and enhances the understanding of the individual's physiology [27,37,47]. It should be emphasized, though, that contextualizing HR in unlabelled data requires navigating many intricacies, as investigators need to account for external factors that can influence HR, such as emotional stress, illness, heat, or medications that can potentially lead to invalid results.

In addition, a key observation arising from studying the multimodal features is the inverse relationship between VO_2max and HR at a given physical activity, which conforms to what is seen during the submaximal tests [5,29,33]. This observation reinforces the concept of estimating fitness from free-living activity, as even in the absence of controlled settings, behaviors approximating submaximal laboratory conditions will spontaneously occur [47]. Neshitov et al [29] demonstrated this inverse relationship, with the slope of the HR-over-cadence regression line being lower for participants with high than those with low VO_2max, and it was mainly noticeable between the 60 - 100 steps per minute exercise effort. Interestingly, Bonomi et al [16] highlighted the need for activity-specific prediction equations, showcasing models that combined energy expenditure and HR based on different activity types. Ultimately, tailoring predictive models to account for specific activity patterns and physiological responses enhances the accuracy of CRF predictions.

Challenges

Despite their potential, free-living data present some intrinsic statistical and computational challenges [48]. Using wearables in out-of-hospital, free-living settings often results in lower-quality, noisy data that require heavy filtering and preprocessing to become usable. Vigorous human motion can disturb on-body sensors and easily corrupt cardiac and accelerometer signals [40]. Aside from noise, missing data can also prevent meaningful features from being extracted. As reported in the "Results" section, preprocessing techniques are an essential step in the data-mining process to ensure that only reliable data points contribute to predictions. Another challenge lies in the precise physical activity detection in unlabeled data. Owing to the diverse nature of daily living, activity patterns overlap, and assumptions are occasionally made on how certain patterns in the data correlate with physical activities [5,49]. Furthermore, the abundance of sedentary data often leads to a data imbalance bias when low-intensity activities are overrepresented, leading to inaccurate estimations [5,16]. Consequently, due to the novelty of the task and the challenges outlined, there is currently no consensus on a specific approach and model that is most suited for free-living data.

Regression and ML Approaches

Interestingly, in the subgroup analysis, regression-based models appeared to perform slightly better than ML approaches. Much research examining the 2 approaches has repeatedly demonstrated comparable performance between regression-based and ML approaches, but this is not universal [38,39]. Nonetheless, this finding from our review warrants attention, as it likely stems from issues such as reporting bias and overfitting rather than genuine superiority. Results are influenced by the notable disparity in sample sizes, with 1 study driving the pooled estimate in ML modeling studies [27]. Finally, the limited external validation in ML suggests that the robustness and interpretability of simpler models can, in some cases, outweigh the complexity when appropriate validation has not been considered.

Limitations

This systematic review showed that free-living data can be valuable for CRF prediction and may prove a useful alternative in a variety of clinical settings. However, some limitations merit attention. First, although we observed promising preliminary agreement between VO_2max estimates and predictions, we need to acknowledge that, although we chose the correlation coefficient as the primary effect size for the meta-correlation analysis for its availability, it is not an accuracy metric and therefore does not imply that predictions are close in absolute terms. Further, as in certain instances, conversion of the R^2 was applied, this may have artificially inflated perceived predictive ability and, as such, influenced the overall result. Error-based metrics such as root-mean-square error or SEE would better capture accuracy, which is particularly relevant in clinical settings, but these measures were not consistently reported.

Second, there was significant heterogeneity and variance among the included studies, which varied in trial design, sample size, wearable device used, and modeling approach. Understandably, this limits the generalizability of our results, and the pooled estimate needs to be interpreted cautiously. However, in contrast to the synthesis of randomized trials, heterogeneity is frequently noted in meta-analyses of predictive modeling studies, mainly due to the disparity of eligible study designs or models [50].

Some studies used resting HR as a feature in their models, limiting monitoring to daytime periods only, excluding nocturnal HR data [31,32,36,37]. Research, however, demonstrates that using nighttime data can yield closer estimations to true testing values when it comes to resting HR [51]. As such, using daytime-derived resting HR may have implications for model performance and potentially lead to erroneous results. Godkin et al [51] underline the lack of standardization and considerable shortcomings among the criteria and methods used to estimate

resting HR. Since daytime HR can be affected by numerous behavioral, psychological, and environmental factors, we advocate for continuous monitoring of free-living data that captures both activity and rest phases for a more stable profile of HR distributions.

Considerations should also be made when interpreting the outputs of the presented models, as the papers reporting the highest correlations between predicted and actual measurements share several common features. Notably, several studies presented outputs from regression models without using unseen data for validation. Directly comparing these outputs against models validated on unseen data likely overestimates the model's true predictive ability. In such cases, high performance may reflect only how well the data fit the model, not its ability to generalize. Therefore, the lack of external validation may contribute to inflation in the aggregated meta-correlation analysis. Under such conditions, a single pooled correlation does not necessarily reflect a uniform level of accuracy and should certainly not be viewed as evidence of clinical readiness. Instead, it demonstrates the potential for future research that a strong association may be achieved.

Another limitation concerning the applicability and validation of the reported models is the selection bias, as most trials recruited young, healthy volunteers, making models less applicable to patient populations. We found that models predict VO_2max with an average 9%, which, arguably, may be clinically relevant in borderline cases—for instance, during preoperative risk assessment of frail individuals. Consequently, no direct conclusions can be drawn for clinical decision-making, and future research should focus more on medical settings to assess the effect of patient-specific factors, such as regular medication and comorbidities, in model training. Interestingly, although there were several different devices used in the included trials, no study explored the practicalities of monitoring patients remotely to collect the free-living data, which could explain the quality issues of noise and missingness that these datasets exhibit.

Implications for Real-World Use

The absence of a shared methodological framework across studies remains a major barrier to translation. As stated previously in the "Limitations" subheading, heterogeneity in this review was high, with studies using distinct device types, signal processing strategies, and validation tactics, often with insufficient external testing. Consequently, this limits confidence in generalizability and reproducibility, making direct comparisons difficult. From a clinical perspective, a model optimized for one setting may fail to transfer effectively in different patient groups, sensors, or wear patterns [52].

For clinical adoption to become feasible, several credibility issues need to be addressed. Pitfalls such as model overfitting, lack of standardized analytical pipelines, and limited evidence that performance is stable under real-world conditions (device updates, medication effects on HR) present significant challenges. Similar challenges have been identified in studies of prediction modeling for CVDs, highlighting the need for independent tools to assess replicability and external validation [52]. Therefore, until uniform data handling and transparent external validation become routine, results from research in this field should be considered promising but not ready for clinical application.

Future Considerations

Despite the challenges and limitations identified in this review, several models reported here should not be overlooked in this expanding research field. Future research should aim to streamline the deployment of wearable devices in out-of-hospital settings and educate and support patients and clinical teams. Furthermore, given the increasing influence of the wearable industry in health care, it is essential for such predictive models to undergo rigorous validation before being fully integrated into clinical practice. Establishing consensus on feature extraction, validation, and reporting guided by frameworks such as TRIPOD-AI and recent calls for transparency in wearable research (INTERLIVE) are recommended for future research to yield reproducible and clinically useful results [21,53].

For public health systems, regulatory frameworks regarding the digital storage, privacy, and security of the vast amount of patient-generated data should be considered early. With the myriad of wearables available, it is important that feasibility work is undertaken to set standards for a reliable and accurate technology, helping avoid a repetitive cycle of temporary models being developed that cannot be extrapolated to clinical contexts or used in clinical practice. Finally, a cost-effective analysis will determine the viability of these remote monitoring systems, ensuring they offer a sustainable solution for patients and health care systems.

Conclusion

This work explores a novel concept for CRF estimation from unsupervised free-living patient data. Contrary to the current gold-standard CPET, which is a snapshot of the individual's functional capacity, wearable health monitoring in free-living conditions generates rich datasets that can be exploited to train models for fitness estimation. Several models are discussed in this paper, with studies applying ML to mine raw data and enhance accuracy.

The combined results from this review show promise, with good preliminary agreement between predictions and measured values. However, no firm conclusions can be drawn for clinical implementation due to the heterogeneity of the studies and the lack of external validation. Nonetheless, continuous data streams appear to be a valuable resource for ML methods to shed light on human behavior and health, leading to a step change in how we measure and monitor CRF, ultimately aiming to improve health outcomes.

Acknowledgments

This study was supported with in-kind assistance from the National Institute for Health and Care Research's HealthTech Research Centre in Accelerated Surgical Care (HRC-ASC) and Leeds BRC.

Funding

AD is funded through a Leeds Hospitals Charity Research Fellowship (A2001580 – Making Surgery Safer for Patients With Bowel Cancer) and a Bowel Research UK Grant (BRUK_SG_24012). ABS is funded through UK Research and Innovation. This research was funded by the Engineering and Physical Sciences Research Council (grant no EP/S024336/1).

Data Availability

This is a systematic review, and data were extracted from the included studies and presented in the tables above. No additional data are applicable in this instance.

Authors' Contributions

AD conceptualized the study and conducted the literature search. Data extraction was performed by AD, ABS, and MRK. AD and DW carried out the analysis. The original draft was prepared by AD, with review and editing contributed by ABS, DW, DG, JT, and DGJ. Supervision was provided by JT and DGJ.

Conflicts of Interest

None declared.

Multimedia Appendix 1 [DOCX File, 17 KB - mhealth_v14i1e69996_app1.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 276 KB - mhealth_v14i1e69996_app2.docx]

References

1. Mandsager K, Harb S, Cremer P, Phelan D, Nissen SE, Jaber W. Association of cardiorespiratory fitness with long-term mortality among adults undergoing exercise treadmill testing. JAMA Netw Open 2018 Oct 5;1(6):e183605. [doi: 10.1001/jamanetworkopen.2018.3605]
2. Plasqui G, Westerterp KR. Accelerometry and heart rate as a measure of physical fitness: proof of concept. Med Sci Sports Exerc 2005 May;37(5):872-876. [doi: 10.1249/01.mss.0000161805.61893.c0] [Medline: 15870644]
3. Lang JJ, Prince SA, Merucci K, et al. Cardiorespiratory fitness is a strong and consistent predictor of morbidity and mortality among adults: an overview of meta-analyses representing over 20.9 million observations from 199 unique cohort studies. Br J Sports Med 2024 May;58(10):556-566. [doi: 10.1136/bjsports-2023-107849]
4. Novoa NM, Varela G, Jiménez MF, Ramos J. Value of the average basal daily walked distance measured using a pedometer to predict maximum oxygen consumption per minute in patients undergoing lung resection. Eur J Cardiothorac Surg 2011 May;39(5):756-762. [doi: 10.1016/j.ejcts.2010.08.025] [Medline: 21146419]
5. Altini M, Casale P, Penders J, Ten Velde G, Plasqui G, Amft O. Cardiorespiratory fitness estimation using wearable sensors: Laboratory and free-living analysis of context-specific submaximal heart rates. J Appl Physiol (1985) 2016 May 1;120(9):1082-1096. [doi: 10.1152/japplphysiol.00519.2015] [Medline: 26940653]
6. Laveneziana P, Di Paolo M, Palange P. The clinical value of cardiopulmonary exercise testing in the modern era. Eur Respir Rev 2021 Mar 31;30(159):200187. [doi: 10.1183/16000617.0187-2020]
7. Jones L, Tan L, Carey-Jones S, et al. Can wearable technology be used to approximate cardiopulmonary exercise testing metrics? Perioper Med 2021 Dec;10(1). [doi: 10.1186/s13741-021-00180-w]
8. Pritchard A, Burns P, Correia J, et al. ARTP statement on cardiopulmonary exercise testing 2021. BMJ Open Resp Res 2021 Nov;8(1):e001121. [doi: 10.1136/bmjresp-2021-001121]
9. Noonan V, Dean E. Submaximal Exercise Testing: Clinical Application and Interpretation. Phys Ther 2000 Aug 1;80(8):782-807. [doi: 10.1093/ptj/80.8.782]
10. Glaab T, Taube C. Practical guide to cardiopulmonary exercise testing in adults. Respir Res 2022 Jan 12;23(1):9. [doi: 10.1186/s12931-021-01895-6] [Medline: 35022059]
11. Altini M, Casale P, Penders J, Amft O. Cardiorespiratory fitness estimation in free-living using wearable sensors. Artif Intell Med 2016 Mar;68:37-46. [doi: 10.1016/j.artmed.2016.02.002] [Medline: 26948954]
12. Kwon SB, Ahn JW, Lee SM, et al. Estimating maximal oxygen uptake from daily activity data measured by a watch-type fitness tracker: Cross-sectional study. JMIR Mhealth Uhealth 2019;7(6):e13327. [doi: 10.2196/13327]

13. Gulati M, McBride PE. Functional capacity and cardiovascular assessment: Submaximal exercise testing and hidden candidates for pharmacologic stress. Am J Cardiol 2005 Oct;96(8):11-19. [doi: 10.1016/j.amjcard.2005.06.016]
14. Bennett H, Parfitt G, Davison K, Eston R. Validity of submaximal step tests to estimate maximal oxygen uptake in healthy adults. Sports Med 2016 May;46(5):737-750. [doi: 10.1007/s40279-015-0445-1]
15. Tudor-Locke CE, Myers AM. Challenges and opportunities for measuring physical activity in sedentary adults. Sports Med 2001;31(2):91-100. [doi: 10.2165/00007256-200131020-00002]
16. Bonomi AG, Ten Hoor GA, de Morree HM, Plasqui G, Sartor F. Cardiorespiratory fitness estimation from heart rate and body movement in daily life. J Appl Physiol (1985) 2020 Mar 1;128(3):493-500. [doi: 10.1152/japplphysiol.00631.2019] [Medline: 31999530]
17. Ferreira JJ, Fernandes CI, Rammal HG, Veiga PM. Wearable technology and consumer interaction: A systematic review and research agenda. Comput Human Behav 2021 May;118:106710. [doi: 10.1016/j.chb.2021.106710]
18. Molina-Garcia P, Notbohm HL, Schumann M, et al. Validity of estimating the maximal oxygen consumption by consumer wearables: A systematic review with meta-analysis and expert statement of the INTERLIVE network. Sports Med 2022 Jul;52(7):1577-1597. [doi: 10.1007/s40279-021-01639-y] [Medline: 35072942]
19. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med 2021 Mar;18(3):e1003583. [doi: 10.1371/journal.pmed.1003583] [Medline: 33780438]
20. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 5;5(1):210. [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
21. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ 2024 Apr 16;385:e078378. [doi: 10.1136/bmj-2023-078378] [Medline: 38626948]
22. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019 Jan 1;170(1):51-58. [doi: 10.7326/M18-1376]
23. R: The R Project for Statistical Computing. R Core Team. 2021. URL: https://www.r-project.org/ [accessed 2024-09-09]
24. Quintana DS. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015;6. [doi: 10.3389/FPSYG.2015.01549/BIBTEX]
25. Daraj LR, AlGhareeb M, Almutawa YM, Trabelsi K, Jahrami H. Systematic review and meta-analysis of the correlation coefficients between nomophobia and anxiety, smartphone addiction, and insomnia symptoms. Healthcare (Basel) 2023 Jul 19;11(14):2066. [doi: 10.3390/healthcare11142066] [Medline: 37510507]
26. Bakbergenuly I, Hoaglin DC, Kulinskaya E. Methods for estimating between-study variance and overall effect in meta-analysis of odds ratios. Res Synth Methods 2020 May;11(3):426-442. [doi: 10.1002/jrsm.1404] [Medline: 32112619]
27. Spathis D, Perez-Pozuelo I, Gonzales TI, et al. Longitudinal cardio-respiratory fitness prediction through wearables in free-living environments. NPJ Digit Med 2022 Dec 1;5(1). [doi: 10.1038/s41746-022-00719-1]
28. Wu Y, Spathis D, Jia H, et al. Turning silver into gold: domain adaptation with noisy labels for wearable cardio-respiratory fitness prediction. . Preprint posted online on Nov 28, 2022. [doi: 10.48550/ARXIV.2211.10475]
29. Neshitov A, Tyapochkin K, Kovaleva M, et al. Estimation of cardiorespiratory fitness using heart rate and step count data. Sci Rep 2023 Sep 22;13(1):15808. [doi: 10.1038/s41598-023-43024-x] [Medline: 37737296]
30. Spathis D, Perez-Pozuelo I, Brage S, Wareham NJ, Mascolo C. Self-supervised transfer learning of physiological representations from free-living wearable data. Presented at: ACM CHIL '21: ACM Conference on Health, Inference, and Learning; Apr 8-10, 2021. [doi: 10.1145/3450439.3451863]
31. Ahn JW, Hwang SH, Yoon C, Lee J, Kim HC, Yoon HJ. Unobtrusive estimation of cardiorespiratory fitness with daily activity in healthy young men. J Korean Med Sci 2017;32(12):1947. [doi: 10.3346/jkms.2017.32.12.1947]
32. Frade MCM, Beltrame T, Gois MDO, et al. Toward characterizing cardiovascular fitness using machine learning based on unobtrusive data. PLOS ONE 2023 Mar 1;18(3):e0282398. [doi: 10.1371/journal.pone.0282398]
33. Plasqui G, Westerterp KR. Accelerometry and heart rate as a measure of physical fitness: cross-validation. Med Sci Sports Exerc 2006 Aug;38(8):1510-1514. [doi: 10.1249/01.mss.0000228942.55152.84] [Medline: 16888467]
34. Cao ZB, Miyatake N, Higuchi M, Ishikawa-Takata K, Miyachi M, Tabata I. Prediction of VO2max with daily step counts for Japanese adult women. Eur J Appl Physiol 2009 Jan;105(2):289-296. [doi: 10.1007/s00421-008-0902-8] [Medline: 18985375]
35. Cao ZB, Miyatake N, Higuchi M, Miyachi M, Ishikawa-Takata K, Tabata I. Predicting VO2max with an objectively measured physical activity in Japanese women. Med Sci Sports Exerc 2010 Jan;42(1):179-186. [doi: 10.1249/MSS.0b013e3181af238d] [Medline: 20010115]
36. Zhang Y, Wang X, Pathiravasan CH, et al. Association of smartwatch-based heart rate and physical activity with cardiorespiratory fitness measures in the community: Cohort study. J Med Internet Res 2024 Jun 13;26(1):e56676. [doi: 10.2196/56676] [Medline: 38870519]
37. Beltrame T, Amelard R, Wong A, Hughson RL. Extracting aerobic system dynamics during unsupervised activities of daily living using wearable sensor machine learning models. J Appl Physiol 2018 Feb 1;124(2):473-481. [doi: 10.1152/japplphysiol.00299.2017]

38. Lindsay T, Westgate K, Wijndaele K, et al. Descriptive epidemiology of physical activity energy expenditure in UK adults (The Fenland study). Int J Behav Nutr Phys Act 2019 Dec 9;16(1):126. [doi: 10.1186/s12966-019-0882-6] [Medline: 31818302]
39. Kannel WB, Kannel C, Paffenbarger RS, Cupples LA. Heart rate and cardiovascular mortality: The Framingham study. Am Heart J 1987 Jun;113(6):1489-1494. [doi: 10.1016/0002-8703(87)90666-1]
40. Syversen A, Dosis A, Jayne D, Zhang Z. Wearable sensors as a preoperative assessment tool: A review. Sensors (Basel) 2024 Jan 12;24(2):482. [doi: 10.3390/s24020482] [Medline: 38257579]
41. Khalid S, Khalil T, Nasreen S. A survey of feature selection and feature extraction techniques in machine learning. 2014 Presented at: 2014 Science and Information Conference (SAI); Aug 27-29, 2014. [doi: 10.1109/SAI.2014.6918213]
42. Kang S, Paul A, Jeon G. Reduction of mixed noise from wearable sensors in human-motion estimation. Computers & Electrical Engineering 2017 Jul;61:287-296. [doi: 10.1016/j.compeleceng.2017.05.030]
43. Kumar A, Dubey P, Ranjan A. Assessment of anxiety in surgical patients: An observational study. Anesth Essays Res 2019;13(3):503. [doi: 10.4103/aer.AER_59_19]
44. Mück JE, Ünal B, Butt H, Yetisen AK. Market and patent analyses of wearables in medicine. Trends Biotechnol 2019 Jun;37(6):563-566. [doi: 10.1016/j.tibtech.2019.02.001]
45. Waterland JL, Ismail H, Granger CL, et al. Prehabilitation in high-risk patients scheduled for major abdominal cancer surgery: a feasibility study. Perioper Med 2022 Aug 23;11(1). [doi: 10.1186/s13741-022-00263-2]
46. Franklin BA, Eijsvogels TMH, Pandey A, Quindry J, Toth PP. Physical activity, cardiorespiratory fitness, and cardiovascular health: A clinical practice statement of the American Society for Preventive Cardiology Part II: Physical activity, cardiorespiratory fitness, minimum and goal intensities for exercise training, prescriptive methods, and special patient populations. American Journal of Preventive Cardiology 2022 Dec;12:100425. [doi: 10.1016/j.ajpc.2022.100425]
47. Hallgrímsson HT, Jankovic F, Althoff T, Foschini L. Learning individualized cardiovascular responses from large-scale wearable sensors data. . Preprint posted online on Dec 4, 2018 URL: http://arxiv.org/abs/1812.01696 [accessed 2024-06-04]
48. Ferguson M, Shulman M. Cardiopulmonary exercise testing and other tests of functional capacity. Curr Anesthesiol Rep 2022;12(1):26-33. [doi: 10.1007/s40140-021-00499-6] [Medline: 34840532]
49. Wang Z, Zhang Q, Lan K, et al. Enhancing instantaneous oxygen uptake estimation by non-linear model using cardio-pulmonary physiological and motion signals. Front Physiol 2022;13:Wang. [doi: 10.3389/fphys.2022.897412]
50. Ahmed I, Debray TP, Moons KG, Riley RD. Developing and validating risk prediction models in an individual participant data meta-analysis. BMC Med Res Methodol 2014 Dec;14(1). [doi: 10.1186/1471-2288-14-3]
51. Godkin FE, Van Ooteghem K, Beyer KB, et al. Measuring resting heart rate during daily life using wearable technology: Examining the impact of behavioral context and methodological criteria. Digit Health 2025 May;11. [doi: 10.1177/20552076251367506]
52. Cai YQ, Gong DX, Tang LY, et al. Pitfalls in developing machine learning models for predicting cardiovascular diseases: Challenge and solutions. J Med Internet Res 2024;26:e47645. [doi: 10.2196/47645]
53. Schumann M, Feuerbacher JF, Heinrich L, et al. Using free-living heart rate data as an objective method to assess physical activity: A scoping review and recommendations by the INTERLIVE-network targeting consumer wearables. Sports Med 2025 Feb;55(2):275-300. [doi: 10.1007/s40279-024-02159-1]

Abbreviations

CPET: cardiopulmonary exercise testing
CRF: cardiorespiratory fitness
CVD: cardiovascular disease
ECG: electrocardiogram
HR: heart rate
ML: machine learning
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Study Risk of Bias Assessment Tool
SC: step count
SEE: standard error of estimate
SVM: support vector machine
TRIPOD-AI: Transparent Reporting of a multivariable or machine learning prediction model for Individual Prognosis Or Diagnosis–artificial intelligence
VO_2 max: maximal oxygen uptake

Edited by L Buis; submitted 13.Dec.2024; peer-reviewed by A Martinko, B Xiang, S Tedesco, S Jiang; revised version received 26.Oct.2025; accepted 17.Dec.2025; published 27.Jan.2026.

Please cite as:
Dosis A, Syversen AB, Kowal MR, Grant D, Tiernan J, Wong D, Jayne DG
Exploiting Unsupervised Free-Living Data for Cardiorespiratory Fitness Estimation: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e69996
URL: https://mhealth.jmir.org/2026/1/e69996
doi:10.2196/69996

Digital Health Interventions to Support Chronic Disease Management: Systematic Scoping Review

Abdullah Al Mahmud[1], PhD; Shane Joachim[2], MSc; Prem Prakash Jayaraman[3], PhD; Caitlin Learmonth[4], MSc; Shivani Tyagi[5], PhD; Abdur Rahim Mohammad Forkan[2], PhD; Muhammad Shuakat[6], PhD; Nilmini Wickramasinghe[7], PhD; Jack Wheeler[8], MSc; Stephanie Best[9], PhD; Alison Trainer[8], PhD

[1]Centre for Design Innovation, Department of Architectural and Industrial Design, Swinburne University of Technology, John St., Hawthorn, Melbourne, Australia

[2]Department of Computing Technologies, Swinburne University of Technology, Melbourne, Australia

[3]School of Science, Computing and Engineering Technologies, Factory of the Future and Digital Innovation Lab, Swinburne University of Technology, Melbourne, Australia

[4]Swinburne University of Technology, Melbourne, Australia

[5]Department of Communication Design, School of Design and Architecture, Swinburne University of Technology, Melbourne, Australia

[6]La Trobe University, Melbourne, Australia

[7]School of Computing, Engineering & Mathematical Sciences, La Trobe University, Melbourne, Australia

[8]Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre, Melbourne, Australia

[9]School of Health Sciences, University of Melbourne, Melbourne, Australia

Corresponding Author:
Abdullah Al Mahmud, PhD
Centre for Design Innovation, Department of Architectural and Industrial Design, Swinburne University of Technology, John St., Hawthorn, Melbourne, Australia

Abstract

Background: Health interventions delivered by digital platforms are gaining popularity and are evolving to address the needs of patients with chronic diseases. The heterogeneity of chronic diseases requires that digital health platforms vary in their approaches to chronic disease management.

Objective: This review aimed to explore the characteristics of digital health platforms and the corresponding digital interventions developed to support patients with chronic diseases. This includes those platforms' design, development, and the metrics by which any incremental benefits they provide are assessed.

Methods: We searched electronic databases including Scopus, Web of Science, PsycINFO, IEEE Xplore, MEDLINE, and Embase. Relevant articles published from January 2013 to November 2024 were extracted. Extracted data were then synthesized using qualitative content analysis and presented in narrative form with relevant tables.

Results: In total, we identified 69 digital health platforms supporting the management of 20 chronic diseases. Most platforms were mobile apps (n=22) or a combination of web and mobile apps (n=15). Most of the platforms (n=44) were tailored to support self-management of chronic diseases. These platforms also provided a web-based portal where health care providers could review and manage the information recorded by patients. In 77% (53/69) of the studies, patients reported that the digital interventions delivered by the platform improved their quality of life, their health, and their ability to self-manage their chronic diseases. In addition, health care providers reported positive outcomes, including improved clinical utility and patient communication. While short-term health outcomes of the digital health interventions were largely positive, long-term health outcomes remain unknown. This was because most of the studies were short-term pilots and often formative in nature (n=42). Many had limited sample sizes, limited participant uptake of the digital platforms, and technical issues. In many cases, further personalization of platforms was required to meet patients' self-management needs.

Conclusions: Digital health interventions can be beneficial in the management of chronic disease. The adoption of digital interventions in combination with regular clinical care can improve health outcomes, support self-management, and enhance communication between patients and health care providers. However, long-term user engagement is the major barrier to their long-term success. High dropout rates, often resulting from a lack of motivation or technical issues, testify to the need for adaptive, low-burden interventions that function seamlessly in users' daily lives. Adopting user-centered and co-design approaches that engage both clinicians and patients in designing digital health platforms may enhance the usability and uptake of such platforms.

(JMIR Mhealth Uhealth 2026;14:e63742) doi:10.2196/63742

KEYWORDS

mHealth; chronic disease; digital technology; co-design; user-centered design; user centered; chronic disease management; support; scoping review; design; development; digital health platform; qualitative content analysis; web-based; self-management; quality of life; clinical utility; patient communication

Introduction

Background

Chronic diseases are defined as long-lasting conditions (ie, 1 y or more) and require ongoing care and account for 41 million (74%) deaths worldwide [1-3]. They can cause disability that affects the quality of life and reduces life expectancy. Patients with chronic diseases may face several challenges when managing their condition, such as (1) conflicting knowledge about the disease or how to manage it, (2) access to care, and (3) communication with health care providers [4]. Chronic diseases pose a significant burden on health care systems, families, and caregivers [5]. Therefore, the prevention and management of chronic diseases has become a global priority, as the prevalence of those diseases can undermine social and economic development.

Digital health interventions use technology to deliver health care services or treatments and facilitate knowledge exchange [6]. These interventions are designed to enhance the quality of patient care by capturing and conveying information in a digital format. Digital health interventions may involve electronic medical records (EMRs), mobile apps or web applications, and wearable sensors such as Fitbit (Google). Technologies such as digital health platforms are gaining increased use in managing chronic diseases [7]. The proliferation of mobile apps and the ubiquitous nature of information technology have fueled the development of platforms that support the management of chronic disease [8,9]. The digital health technologies discussed in this study include mobile apps, web applications, electronic health records (EHRs), EMRs, wearable devices, and telehealth services [10]. We have used digital platforms as an umbrella term throughout the paper to denote these technologies.

With the rapid growth of digital platforms for chronic disease management, a systematic synthesis of these platforms is needed to inform effective and efficient care. Recent studies have investigated the framework for managing chronic diseases [11] and the potential of technology adoption [12]. To our knowledge, no studies have examined how digital platforms support the management of chronic diseases. Due to the heterogeneity of digital platforms, it is vital to investigate the types of digital platforms that are available and the usability and acceptability of these platforms. In addition, it is important to investigate the processes that led to their design and development, as well as the metrics used to assess their benefits. In this study, digital platforms are characterized as assortments of web-based and mobile applications and related technologies that are used to deliver health care services [13].

Aims

This review aims to explore the characteristics of digital health platforms and corresponding digital health interventions that support patients with chronic disease. This exploration will include those technologies' design, their development processes, and the metrics by which their incremental and long-term benefits have been assessed.

Methods

Overview

This review adopts the scoping review methodology proposed by Arksey and O'Malley [14] because we are interested in identifying and mapping emerging evidence [15]. PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist [16] was used throughout the review to ensure adherence (Checklist 1). However, there are some differences between the registered protocol and this paper regarding database searching. Furthermore, 2 additional databases (Medline and Embase) have been searched, and the search period has been extended to November 2024 for all the databases.

Step 1: Identifying the Research Questions

First, what are the characteristics of digital platforms that support the management of chronic diseases, including self-management and provider-led management?

Second, what principles and theoretical frameworks have been used to design or co-design these platforms?

Third, how were these platforms evaluated for clinical utility?

Finally, what is the effectiveness of those platforms?

Step 2: Search Strategy

The search terms used for the literature search are "chronic disease" OR "chronic illness," OR "long-term conditions," OR "chronic conditions" AND "Digital" OR "mHealth" OR "App" AND "management."

We did not include the exact term "self-management" in our search, nor the standard indexing terms that databases use for that idea (eg, PubMed's MeSH heading Self Care and Embase's Emtree term self-management). To reduce the chance of missing papers, we implemented citation chasing, that is, for each included study, we checked its reference list and looked up newer papers that cite it.

Initially, databases such as Scopus, Web of Science, PsycINFO, and IEEE Xplore were searched from January 2013 to 30 November 2022. Later, an additional search was conducted on these databases from November 2022 to November 2024. Furthermore, 2 additional databases, MEDLINE and Embase, were searched from January 2013 to November 2024 (Refer to Multimedia Appendix 1 for the search outcomes). Articles retrieved were imported into Covidence software (Veritas Health Innovation Ltd) [17], and duplicated items were automatically removed.

Step 3: Study Selection

The following inclusion and exclusion criteria were used.

Inclusion criteria

First, the study must be a peer-reviewed journal article and present primary data. Second, it should be published within the last 12 years (January 2013 to November 2024). We chose 2013 as the starting year to reflect a critical turning point in the evolution of digital health technologies. Around this time, the widespread use of smartphones and mobile apps, breakthroughs in wearable sensors, and more prevalent usage of EHRs began to transform chronic disease management [18,19]. Third, it should be available in English. Finally, it should involve digital applications in the context of chronic disease management.

Exclusion criteria

The study was excluded if (1) it was a review article or opinion piece, or (2) hypothetical use of digital technology was found in it.

Initially, titles and abstracts were reviewed against the selection criteria and were marked as "include," "exclude," or "uncertain." Two reviewers (AAM and MS) conducted the screening independently, and regular discussions with the research team were undertaken to resolve any discrepancies and to fine-tune selection criteria. This screening and discussion process continued until we reached a consensus. Subsequently, for the included studies, a full-text review was carried out (AAM and MS) against the selection criteria following the same screening procedure.

Step 4: Data Extraction and Charting the Data

Two authors (AAM and MS) developed a data charting form to identify relevant information to extract from the included studies. Using this form, the following data were extracted: study citation, publication type, authors, study location, study year, user acceptability of the digital interventions or digital platforms, and outcome of the study (quantitative results, qualitative themes, recommendations, key learnings, and limitations). AAM and MS charted the extracted information.

Step 5: Collating, Summarizing, and Reporting the Results

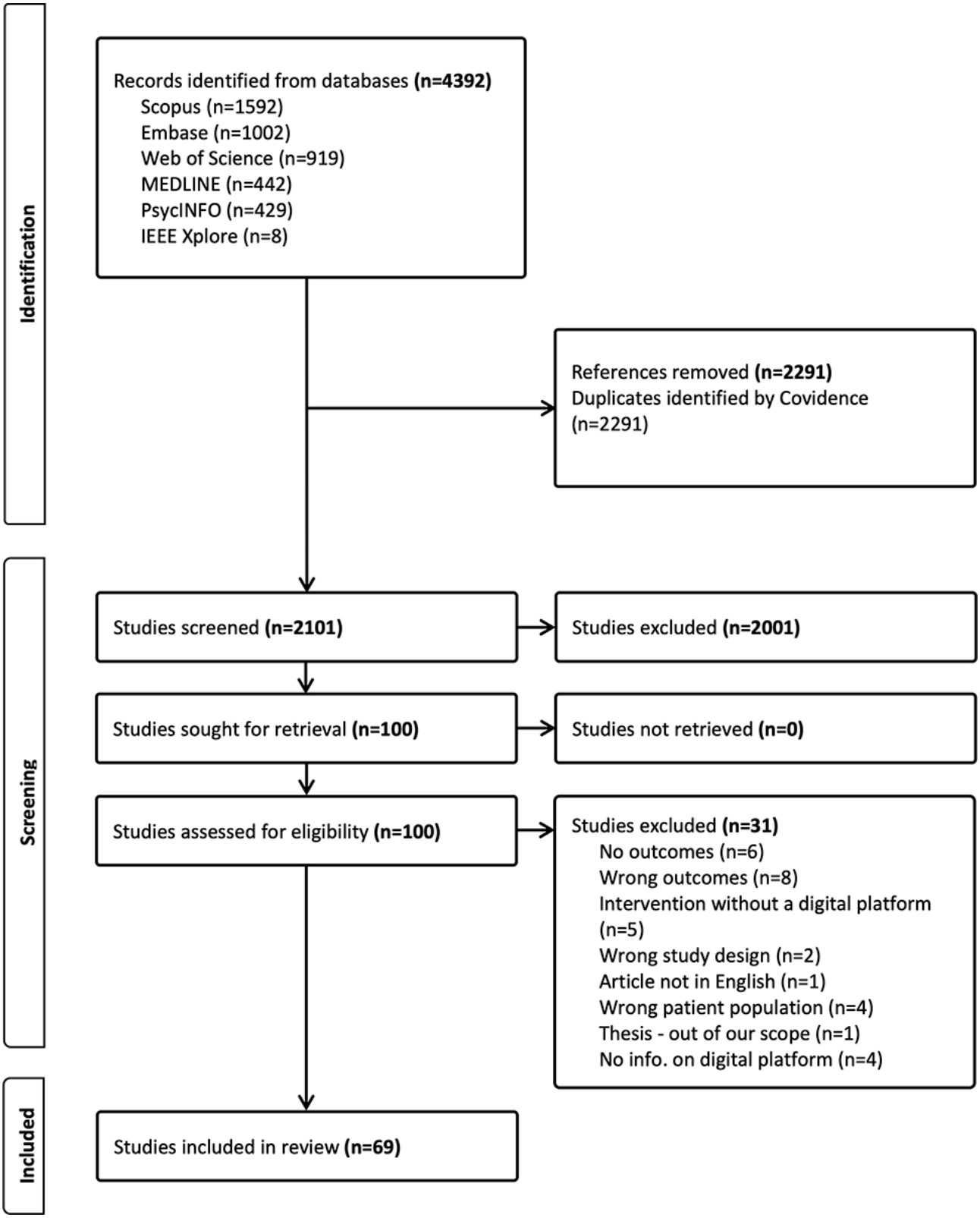
First, the extracted data were analyzed using descriptive statistics (eg, frequencies). This provided numerical summaries of (1) digital platforms and their characteristics, (2) chronic diseases, (3) platform design principles, and (4) outcomes. These details were presented using tables, charts, and graphs, followed by a summary. Second, 2 authors (AAM and CL) independently analyzed the extracted data thematically to identify themes. Results from the 2 reviewers' thematic analysis were combined to select the final collection of themes.

Results

Overview

A total of 4392 studies were identified from the 4 databases. Of these, 2291 were duplicates, leaving 2101 to be screened. In total, 2001 studies were excluded during the title and abstract screening process, and 100 were assessed for eligibility. Of these, 69 studies [20-88] met our eligibility criteria and were included for review. Refer to Figure 1 for the article selection process.

Figure 1. PRISMA diagram showing the article selection process.



Characteristics of the Included Studies

In total, 69 studies were conducted in 26 countries. Of the total, 15 studies (22%) were conducted in the United States [22,25-27,30,31,37,38,49,50,52,55,61,66,85]. Canada had 5 studies (7%) [38,49,50,76,84], and China had 10 studies (15%) [24,32,41,43,58,63,65,69,79,81], while 4 studies (5%) [28,59,67,83] were conducted in Spain. The most common chronic disease reported was type 2 diabetes (n=17, 20%), followed by heart failure (n=11, 13%), chronic obstructive pulmonary disease (COPD; n=9, 11%), and hypertension (n=9, 11%) (Multimedia Appendix 2). In total, 27 studies (39%) [21,22,24,27-29,31,33,34,37,41,42,45,49,50,52,54,55,59,60-68,69] are reported as randomized controlled trials (RCTs), and the

remaining 42 studies (61%) are formative studies [20,23,25,26,30,32,35,36,38-40,43,44,46-48,51,53,56-58,61,62,70-88]. Table 1 summarizes the aims of the digital platforms, the platform types, and the study settings.

Table . Summary of the platforms.

Platform characteristics	Total, n (%)
Aim of the digital platform (n=77)	
Self-management[a]	44 (57)
Behavior changes	17 (22)
Communication with health care providers	16 (20)
Types of digital platforms (n=69)	
Web-based application	15 (22)
Mobile app	33 (48)
Combination of web-based application and mobile app (ie, multimodal)	14 (20)
Other[b] (SMS text messaging)	7 (10)
Wearable device (in combination with web app, smartphone app, or both)	16 (23)
Study setting (n=78)	
Hospital or primary care setting	34 (44)
Home	36 (46)
Online Community	8 (10)

[a]Includes self-monitoring of symptoms, medication, physical activity, etc.

[b]One intervention used telehealth and SMS text messaging.

Characteristics of the Study Participants

All the included studies had adult participants (aged >18 y). Participants from various age groups were included in the studies (30 - 35 y: n=2, 40 - 45 y: n=1, 45 - 50 y: n=1, 60 y or older: n=19, and 75 - 80 y: n=1). Several studies (n=15) did not specify the mean age or age range. Most of the studies are dominated by female participants, especially studies related to the management of chronic diseases such as asthma (eg, 80 women vs 26 men [20]), hypertension (eg, 49/67, 73% female [21]), diabetes, and rheumatoid arthritis. Some studies target specific gender groups to address health disparities or disease prevalence, such as Black women with hypertension [22]. Most chronic conditions are higher in females, which may be a reason for overrepresentation. Some studies, like those dealing with heart failure (eg, N=25, 100% male participants [23]), atrial fibrillation (eg, 59/96, 61.5% male [24]), and other cardiovascular disorders (eg, 48/79, 61% male [54]), might have a greater proportion of males. In some instances, these conditions may be more prevalent in men within specific age groups. Most male-dominated studies occur in areas such as cardiovascular health, where men are more prone to certain conditions.

In total, 6 studies [40,54,57,58,71,89] mentioned the literacy levels of participants in the participant eligibility criteria. Only 1 study [26] mentioned digital literacy (ie, technical and cognitive abilities to use information and communication technologies [89]), defined as "acceptable literacy level to read and write with a smartphone," in the inclusion criteria. None of the studies measured the participants' digital literacy level as part of the intervention. Most studies (n=19; 61%) had predominantly male participants (66%). Where literacies were mentioned, the most reported literacy was linguistic (English language literacy, n=2 out of 6), followed by health literacy [90], defined as the capability to process and understand health information (n=2 out of 6).

Characteristics of the Digital Interventions

The digital interventions identified in the literature targeted a range of chronic diseases and thus had a variety of operational aims. Multimedia Appendix 3 provides a summary of the digital intervention strategies used in the studies. The definition of each digital intervention is extracted from the National Institute for Health and Care Excellence (NICE) Framework [91]. An analysis of the digital intervention strategies based on the NICE Evidence Standards Framework for digital health [1] found self-management to be the most prevalent strategy, appearing in 68 studies (99%) [20-46,49-86,88]. Collaborative care (ie, self-management plus provider-led management) followed with 43 occurrences (39%) [20-25,27,28,32-37,39-42,44,51-53,56,57,60-63,65,67,69,70,72,74,77-83,85,87], while information and education were present in 27 studies (39%) [20-24,26,33,37,40,45,46,57,59,63,64,66,72-79,81,84,85]. Personal health record systems were identified in 25 papers (62%), digital therapeutics in 15 papers (22%) [22,23,25,28,35,44,45,51,59,60,63,66,72,73,77], clinical decision support systems in 8 papers (12%) [26,28,44,48,59,65,70,72], and active monitoring in 1 paper (2%) [30].

Self-management frequently co-occurs with collaborative care (47 times), reflecting strong integration between patient-driven health management interventions and the need for clinical support (ie, provider-led management). Information and education (31 times) and personal health record systems (25 times) also frequently co-occur with self-management, underscoring the importance of providing users with relevant knowledge and real-time data to make informed health decisions. This, in turn, enhances self-care practices and promotes better health outcomes. Additionally, digital therapeutics (15 times) and clinical decision support systems (8 times) demonstrate strong associations, emphasizing their role in remote treatment guidance and data-driven clinical decision-making. These findings highlight how personalized digital interventions, combined with real-time insights, can enhance both patient engagement and clinical oversight.

Features of the Digital Platforms and Behavior Change Techniques Offered

Overview

The most common features by far are self-monitoring and tracking (65/69 studies, 94%), showing the trend of users' empowerment for active management (Table 2) of chronic diseases with support from health care providers. Medication reminders (59/69 studies, 86%) and behavioral support features (48/69 sources, 70%) are essential in helping patients stay committed to their treatment plans, with alerts (59/69 studies, 86%) and gamification using points, badges, or levels (8/69 studies, 12%) playing key roles in engagement. Most of the platforms are also targeted at improving user awareness and engagement by educating the users about better health practices (55/69 sources, 80%) and communicating with health care providers (45/69 studies, 72%). Such features ensure that patients are supported not only through technology but also by and through human contact.

Table . Summary of the most common features across the 69 studies, including their frequencies.

Category and feature	Frequency
Mobile app (n=32)	
Self-monitoring	15
Patient education	10
Reminders	18
Data tracking	14
Communication (between patient and health care providers)	12
Personalized feedback	8
Medication adherence	7
Activity tracking	9
Health reporting (eg, symptom, weight, and blood pressure)	6
Push notifications	7
Web-based application (n=15)	
Health data management	8
Patient education	5
Health reporting	6
Care team communication	4
Goal setting	3
Progress tracking	4
Data visualization	2
Document management (eg, laboratory results and prescriptions)	2
Personalized feedback	1
Combination of web-based application and mobile app (multimodal; n=16)	
Health data integration	7
Communication (between patient and health care providers)	5
Health tracking	6
Goal setting	5
Personalized feedback	4
Reminders and notifications	5
Symptom tracking	4
Educational content	3
Activity tracking	3
Wearable device (in combination with web application or mobile app; n=16)	
Physical activity tracking	9
Health data measurement (eg, heart rate, blood pressure, oxygen saturation, and SpO2[a])	8
Real-time monitoring	7
Data syncing (with mobile app or cloud)	6
Reminder and alert functionality	5
Health reporting	4
Medication reminders	3
Data visualization	3

Category and feature	Frequency
Other (SMS text messaging; n=7)	
Reminders (medication, appointments, etc)	7
Behavioral triggers	5
Education (health tips and guidance)	6
Symptom reporting	4
Daily check-ins or reports	3

[a]SpO2: peripheral oxygen saturation.

Table 2 provides a summary of the features and their frequencies by category. Routine input of health data was facilitated through surveys and questionnaires, using free-text or drop-down menu functionalities. Some apps facilitated image uploads, allowing patients to supply photos of wounds, rashes, or other relevant aspects of their conditions to be assessed by health care providers [27,28]. To support the self-management and tracking of health data, many platforms use tools for self-measurement and reporting of key health indicators, such as blood pressure (BP), glucose levels, and heart rate. Furthermore, 2 platforms embedded these functionalities in motion sensors [29,30]. However, the majority of platforms incorporating self-reporting functionalities did so through Bluetooth-enabled technology, such as smart watches, BP monitors, and scales [31-36]. These either fed data directly to the platforms or provided data for participants to input manually. Refer to Table 3 for the overview of the wearable devices used in the included studies.

Table . Wearable devices used in the included studies (n=16).

Wearable device	Type of wearable	Purpose	Study
Fitbit	Activity tracker or fitness band	General health monitoring, physical activity, and heart rate tracking	Oh et al [34]
Apple Watch	Smartwatch	ECG[a], heart rate monitoring, activity tracking, and general health monitoring	Guo et al [32]
Omron BP Cuff	Blood pressure monitor	Blood pressure tracking and monitoring for hypertension	Evans et al [30]
Dexcom CGM	Continuous glucose monitor	Continuous glucose monitoring for diabetes	Schnall et al [37]
iHealth Pulse Oximeter	Pulse oximeter	Monitoring oxygen levels (SpO2[b]) for respiratory conditions	Guo et al [32]
AliveCor KardiaMobile	ECG monitor	ECG readings for detecting arrhythmias	Gray et al [38]
Masimo Pulse Oximeter	Pulse oximeter	Monitoring oxygen saturation levels (SpO2)	Kryger et al [27]
Apple Watch	Smartwatch	Heart rate monitoring, ECG, and wellness tracking	Burda et al [39]
Fitbit	Activity tracker or fitness band	Step counting, heart rate monitoring, and sleep tracking	Bailey et al [40]
FreeStyle Libre	Continuous glucose monitor	Continuous glucose monitoring for diabetes management	Dorsch et al [31]
Google Glass	Smart glasses	Augmented reality for hands-free monitoring during medical procedures	Zhang et al [41]
Quell	Pain management wearable	Nerve stimulation for chronic pain relief	Cormican and Dowling [88]
Oura Ring	Wearable sleep tracker	Sleep quality tracking and recovery monitoring	Poppe et al [42]
Masimo	Pulse oximeter	Continuous SpO2 monitoring for respiratory health	Ji et al [43]
WHOOP Strap	Wearable fitness tracker	Sleep and recovery tracking for physical performance enhancement	Doyle et al [44]
iHealth Pulse Oximeter	Pulse oximeter	Monitoring oxygen levels in the blood	Ji et al [43]

[a]ECG: electrocardiogram.

[b]SpO2: peripheral oxygen saturation.

Only 5 studies [29,44-47] explicitly mentioned the use of artificial intelligence (AI) in the digital platforms. In these, AI was used to aggregate and analyze patient data across multimodal platforms. This was achieved through machine learning algorithms or recommender systems on platforms that use conversation agents. For example, the Snapcare app (Snapcare Technologies Pvt Ltd) gathered daily activity data over 12 weeks to address chronic back pain, including walking distance and workouts [29]. Notifications were sent to patients based on app usage and physical activity data collected by built-in phone sensors. This information was automatically transferred to a secure server, where machine learning algorithms examined the daily data on physical activity and produced suggestions for the session the following day. Similarly, the ProACT digital platform helps older patients manage their multimorbidity, including diabetes, congestive heart failure, COPD, and chronic heart disease [44]. Health information, including BP, heart rate, blood glucose, pulse oximetry, weight, activity, and sleep readings, is collected using off-the-shelf technologies. ProACT's AI algorithms can gain knowledge from the data to provide more accurate personalized recommendations and highlight a condition that needs attention. Likewise, using an embodied conversational agent named Laura, My Diabetes Coach (The University of Melbourne) provides patients with individualized support, monitoring, and motivational coaching [45]. The algorithms and conversational scripts that direct each person's progress were developed using behavior change theories, and they can accommodate recommendations by a general practitioner. Natural language processing and automated speech recognition are used to enhance the capability of a voice-enabled chatbot in user interactions in French [46]. The Medly voice app (University

Health Network) for heart failure management leverages various AI technologies, including machine learning, natural language processing, and automated speech recognition, which process the speech of users and formulate responses [47].

In total, 14 of the platforms included in the review were multimodal. For example, a platform might function as a combination of web-based and smartphone-based applications or work in combination with a wearable Bluetooth device, such as a smartwatch or BP monitor. Of the total, 8 interventions included some form of wearable technology or measurement device. These ranged from smartwatches and activity trackers [30-32,40,44] to BP monitors [33,34,36,44], glucose monitors [34,44], scales [31], and sleep monitoring devices [44]. Additionally, 15 interventions [26-28,35,37,39,44,47,48,50,53,58,61,62,71] were designed to work in conjunction with in-person consultations. These included functionalities that were integrated with routine clinical visits to monitor appointments, provide technical support, and review progress. Digital platforms in 10 studies supported multimorbidity [27,34,35,38,44,48-52].

Behavior change was broadly reported to be a multistep process requiring numerous complementary features. Tailoring features to the specific needs of participants based on both demographics and disease type was shown to be important in achieving behavior change [32,48,53]. Although not all of the studies explicitly used theoretical frameworks for behavioral change, many of the platforms are underpinned by theories intended to enable behavior change. Theories applied include social cognitive theory, which is detailed in the work of Dale et al [54], and the transtheoretical model, which is elaborated upon by Salari et al [53]. Furthermore, Sittig et al [55] explore the Fogg Behavior Model, and both cognitive theory and self-efficacy theory are examined by Park et al [60].

A range of methods was used to deliver tailored behavior change features to facilitate user engagement and improved health outcomes. Reminders, goal-setting, and motivational messages were widely used in 5 studies [38,39,44,56,88]. Those features aimed to motivate patients to engage with their health interventions and help them stay on track with their treatment plans. Behavioral trigger messaging was used to promote engagement, motivation, and self-belief. It was also used to provide reinforcement to participants based on behavior change theories [55]. Similarly, theory-based approaches specific to tailored SMS text messaging were used to improve personalization and increase the acceptability of the health interventions [32,53].

Gamification and learning material were incorporated in some studies for the optimization of motivation and to increase users' knowledge of managing their conditions. Gamification, either through rewards or challenges, stimulated usage of the app, while educational information [26,33,40,57] provided valuable information to allow patients to better understand their conditions and make informed decisions on their treatment. Furthermore, 5 platforms used gamification techniques to enhance user experience while embedding learning principles [29,42,45,58,59]. In-app conversation agents supported by AI [45] were used in a platform to change behavior too.

Furthermore, self-monitoring, tailored feedback, and social support were key features of some platforms [26,35,40,60]. These features fostered engagement and behavior change. Self-monitoring permitted users to monitor their health measurements (ie, BP or blood glucose), while personalized feedback assisted in matching interventions to individual needs and encouraging ownership of one's health. Social support, either through peer interaction or direct contact with health care providers, also facilitated long-term engagement by providing users with emotional encouragement and accountability.

Chronic Disease Management: Patient Self-Management and Provider-Led Care

In analyzing digital platform features, we distinguish 2 complementary approaches: patient self-management and provider-led management. Provider-led management refers to the coordinated, proactive activities delivered by health care services and clinicians to support people living with chronic conditions, typically including comprehensive care planning, risk stratification, guideline-concordant treatment, continuity, between-visit care coordination, and use of clinical information systems for monitoring and quality improvement across settings and providers [92]. In contrast, patient-led self-management encompasses the day-to-day work undertaken by people living with chronic conditions, including symptom monitoring, medication adherence, lifestyle and behavioral changes, decision-making, help-seeking, and the usage of tools and supports, including digital platforms [93].

The vast majority of digital platforms (n=36) were designed to be used at home by patients, and this was usually done in conjunction with their ongoing health care plans. To support at-home interventions, some platforms incorporated automated SMS text messaging, including conversational agents and trigger SMS text messaging [45,55]. The key focus of platforms that were exclusively home-based was self-management through input and self-monitoring of health data. Of the 36 interventions, 13 designed to be used in the home were also implemented in primary care settings, including hospitals. The digital platforms, which contained a patient-facing mobile health (mHealth) app, primarily focused on self-management of the symptoms of a chronic condition. The data collected by an mHealth app is usually passed to a practitioner-facing portal, where the practitioner can view patient data and adjust care plans. In some cases, these digital platforms supported bidirectional communication, allowing the practitioner to directly communicate with patients through the platform [32,35,45,53,60,61]. A few platforms also supported the booking of appointments with relevant health care providers [33,48,49] or sent reminders about upcoming appointments [43].

Because a high number of digital platforms are tailored to support self-management (n=44), the majority of platforms were primarily managed by the patients themselves. However, where the digital platforms were multimodal, consisting of a smartphone app and a web-based application, the smartphone apps tended to be managed by the patients, while the web-based applications were used by health care providers to review and manage patient information through an online portal.

Principles Used in the Design and Development of the Digital Platforms

In total, 25 studies (36%) used some form of co-design, consultative, or user-centered approach to developing the digital platforms. Furthermore, 11 studies focused on platforms that either had already existed or were adapted from existing models. Co-design approaches varied and included expert consultation with health care providers and user-centered iterative approaches, which sought feedback from patients throughout an intervention period.

Most of the platforms were developed by using surveys or focus groups to gather information and determine requirements. Often, these surveys and focus groups lacked the full participation of end users. In some studies, the iterative development approach also included gaining input from caregivers, patients, and related stakeholders. For example, the iMHere 2.0 [35] system was iteratively designed, developed, and evaluated with patients involved at all stages. Other studies that discussed the development of digital platforms did not follow or report on the iterative development process. One platform [62] was not accepted as it failed to meet the needs of the patients. Studies adopting an iterative approach tended to make changes to the digital intervention throughout the testing period in response to feedback, whereas those using professional consultations at the beginning and end of the process did not make changes during the testing period. Platforms primarily seeking to influence behavior change tended to be developed using more theoretically based approaches, such as behavior change models and social cognitive theory [54,55].

Outcomes of Studies Assessing the Digital Platforms

Outcomes of the digital platform assessments reported in the studies are grouped into 2 categories: outcomes of randomized controlled studies and other formative studies.

Outcomes of RCTs

The majority of the included studies demonstrated significant benefits and health outcomes, such as symptom reduction, better disease control, and improved quality of life. A smaller proportion of studies showed no significant effect on the primary outcomes, with improvements in only secondary measures that did not significantly impact the main health indicators. Some trials did not have the expected results, sometimes due to high dropout rates, lack of long-term effects, or problems with patient engagement (Table 4).

Table . Summary of the randomized controlled trial studies (n=27).

Study	Favorable outcomes	No effects	Name of the platform and target disease
Li et al [63]	Significant improvement in disease control at 6 months (*P*=.001).	At 12 months, no significant difference between groups (*P*=.90).	A smart system of disease management; rheumatoid arthritis
Goulding et al [64]	Decreased relapse risk in the low-risk group (*P*=.02), improved depressive symptoms (*P*=.02), and relational quality of life (*P*=.02).	There was no significant improvement in relapse risk for the high-risk group (*P*=.62).	A smartphone-based self-management intervention (LiveWell) app and website; bipolar disorder
Goodman and Locke [21]	Significant improvements in blood pressure were found, with systolic BP[a] decreasing from 140 to 134 mm Hg (*P*=.001) and diastolic BP decreasing from 78 to 74 mm Hg (*P*=.007)	No effect on the usual care group for HbA1c[b] (*P*=.19).	A mobile phone-delivered diabetes intervention; diabetes
Zhang et al [65]	Significant improvements in HbA1c (*P*<.05), diastolic BP (*P*<.05), and fasting plasma glucose (*P*<.05).	No effect on BMI.	A digital health technology to provide shared decision-making–informed dietary intervention; diabetes
Buis et al [66]	Both groups showed significant reductions in systolic BP (*P*<.001).	No significant differences between groups for BP or other outcomes (*P*=.99).	MI-BP, a culturally tailored multibehavior mobile health intervention; hypertension
Zhang et al [41]	Significant improvement in HbA1c (*P*<.001) and ABC[c] control rate (*P*=.025). ABC goals are HbA1c <7%, systolic BP/diastolic BP <140/80 mm Hg, and LDL-C[d] <2.6 mmol/L.	No changes in LDL-C or blood pressure (*P*=.95).	SMARTDiabetes app; diabetes
Tabernero et al [67]	Significant improvement in emotional well-being and self-efficacy for chronic disease management (*P*<.05).	No effect on anxiety and depression levels.	Psychological interventions delivered via mHealth[e] technology; chronic cardiac diseases
Hartch et al [52]	Significant improvements in medication adherence (Cohen *d*=−0.52, *P*=.014) and medication self-efficacy (Cohen *d*=0.43, *P*=.035).	No significant effect on medication knowledge or social support (*P*=.15).	Medisafe app; hypertension, diabetes, and asthma
Xu et al [24]	Significant improvement in anticoagulation knowledge, medication compliance, and patient satisfaction (*P*<.05).	No effect on bleeding or thrombotic events.	Alfalfa app; atrial fibrillation
Abel et al [22]	Significant reduction in systolic BP (*P*<.001), weight (*P*<.001), and physical activity (*P*=.018).	No significant difference in BP control between the treatment and control groups (*P*=.99). No sustained effect on BP control after 6 months.	Chronic disease self-management program; hypertension
Oh et al [34]	Significant reduction in body fat mass (*P*=.04) and HbA1c (*P*=.03) in the integrative mHealth group.	No significant changes in body weight, BMI, blood pressure, or HbA1c.	Integrative mHealth platform; hypertension and diabetes
Sittig et al [55]	Statistically significant improvements in self-efficacy (*P*=.008) and exercise (*P*=.01) in high and mid-users.	No significant differences across groups in overall health measures (ANOVA).	“capABILITY” app; diabetes
Patnaik et al [68]	Significant improvements in weight, BMI, waist circumference, and systolic blood pressure (*P*<.05).	No significant difference in metabolic equivalent (MET) levels (*P*=.54).	A mobile interactive platform: an Android-based application; diabetes
Jia et al [69]	Significant improvement in HbA1c (*P*<.001) and ABC control (*P*=.025).	No effect on hypoglycemia or weight gain (*P*=.95).	Graded the ROADMAP app and a website; diabetes

Study	Favorable outcomes	No effects	Name of the platform and target disease
Lear et al [49]	Fewer hospitalizations and in-hospital days (*P*<.05).	No significant reduction in hospitalization rates (*P*=.12).	An internet-based self-management and symptom monitoring program for diabetes, heart failure, ischemic heart disease, chronic kidney disease, or chronic obstructive pulmonary disease
Gray et al [50]	Participants set meaningful self-management goals.	No significant improvement in quality of life (*P*=.24).	Electronic patient-reported outcome mobile app and portal system; multiple chronic diseases such as arthritis, asthma, and hypertension
Kryger et al [27]	Significant reduction in urinary tract infections (*P*=.03).	No significant change in psychosocial outcomes.	iMHere mHealth system; spinal cord injury
Gong et al [45]	Significant improvement in quality of life (*P*=.04).	No significant change in HbA1c (*P*=.83).	My Diabetes Coach program, an app-based interactive embodied conversational agent; diabetes
Chhabra et al [29]	Significant improvement in disability (*P*<.001).	No significant change in pain.	Snapcare app; low back pain
Puig et al [28]	High patient satisfaction (85.2% find it useful, 91.4% would recommend it).	No significant change in quality of life or clinical outcomes.	+Approp app; HIV
Dale et al [54]	Significant improvements in medication adherence (*P*=.004).	No long-term effects on lifestyle changes (*P*=.13).	An mHealth-delivered comprehensive cardiac rehabilitation program called Text4Heart for coronary heart disease
Dorsch et al [31]	Improvement in Minnesota Living with Heart Failure Questionnaire at 6 weeks (*P*=.04).	No effect on self-reported HF management (*P*=.78).	ManageHF4Life app; heart failure
Velardo et al [33]	High compliance with self-monitoring (96% of symptom diaries completed).	No significant impact on patient outcomes or disease progression.	"Self-management and support programme (EDGE);" COPD[f]
Poppe et al [42]	Improvement in physical activity in some groups (*P*<.05).	Limited improvements in sitting time and moderate physical activity (*P*=.09 to *P*=.07).	MyPlan 2.0 comprises a website and an optional mobile app for diabetes
Schnall et al [37]	Significant improvement in 5 symptoms (anxiety, depression, neuropathy, fever or chills, and weight loss) (*P*<.05).	No effect on other symptoms.	mobile video information provider app; HIV
Morcillo-Muñoz et al [59]	Improvement in catastrophizing, rumination, and quality of life (*P*<.05).	No effect on magnification or satisfaction with health.	NO+Dolor (NO+ Pain) app; chronic pain
Park et al [60]	Improved self-care behavior (*P*=.01) and physical activity.	No improvement in self-efficacy for managing dyspnea. The number of steps per day did not significantly differ at 6 months.	Smartphone app-based self-management program; COPD

[a]BP: blood pressure.

[b]HbA1c: hemoglobin A1c.

[c]ABC: the "ABC" goals for type 2 diabetes management and stands for A1C (a measure of blood sugar), blood pressure, and cholesterol (specifically low-density lipoprotein cholesterol).

[d]LDL-C: low-density lipoprotein cholesterol.

[e]mHealth: mobile health.

[f]COPD: chronic obstructive pulmonary disease.

mHealth technologies are emerging as a promising solution in managing chronic conditions, offering patients a convenient and efficient way to monitor and improve their health. These randomized controlled studies (n=26) that reported favorable outcomes indicate that the impact of the mHealth intervention is evident in chronic disease management [21,22,24,27-29,31,33,34,37,41,42,45,49,50,52,54,55,59,63-69]. The clinical improvements were the more consistent outcome

across many of the included studies. Out of 26 RCTs, 30.8% (n=8) reported favorable improvements in key clinical health metrics, including BP [22,65,66,68], hemoglobin A1c (HbA1c) levels [21,34,41,65,69], weight [22,68], and blood glucose levels [65]. For example, the SMARTDiabetes trial [41] demonstrated better glycemic control in the intervention arm, improving HbA1c, BP, and low-density lipoprotein cholesterol control. These improvements were particularly notable in chronic diseases like diabetes and hypertension, suggesting that mHealth interventions can be effective in helping patients manage these conditions over time.

Beyond symptoms and clinical improvements, many studies found that these mHealth apps could create better self-management; 12 out of 26 studies (46%) demonstrated that patients became more involved in managing their health by tracking symptoms, medication adherence, and healthier behavior changes, such as increased physical activity [22,24,31,33,37,41,42,45,52,55,67,68]. For example, in the bipolar disorder study [64], participants using the smartphone-based self-management intervention showed reduced depressive symptoms and improved relational quality of life. This underlines the fact that mHealth tools are instrumental in improving clinical outcomes and empowering patients to take greater control of their conditions.

A common theme from these studies is increased patient engagement, which is favorable in 14 (54%) of these studies [22,24,31,33,37,41,42,45,52,54,55,63,67,68]. Most participants shared that engaging in mobile apps to track symptoms and remind themselves about their medication and educational content increased patients' interest in their health. In the study conducted on rheumatoid arthritis, it was observed that patients who were exposed to the mobile app had better control of their disease, while the smart system of disease management group [63] had a higher rate of patients with controlled disease than the control group, at 71% versus 64.5%.

Besides engagement, user satisfaction also appeared consistently high across the positive-outcome studies. Specifically, participants expressed appreciation for the ease of use in navigating the interfaces, personalized feedback, and easy access to health care information in studies by Xu et al [24] and Puig et al [28]. For example, in the study on the Alfalfa App [24], there was improved medication adherence among patients, $P<.001$, and a very high satisfaction regarding the app's utility in managing anticoagulation therapy.

Another important insight from the positive studies is the reduction in disease symptoms, reported in 5 of the 26 RCTs (19.2%) [27,31,37,63,64]. Indeed, many studies reported significant improvements in specific symptoms such as pain, depression, anxiety, and fatigue. This was especially true in conditions like HIV, where the mobile video information provider app [37] helped alleviate symptoms, such as neuropathy, anxiety, and depression, while also increasing medication adherence.

mHealth interventions also helped most patients improve their overall quality of life. Many studies have shown evidence for the above fact as the common resultant factor in the case of chronic diseases. Statistical improvements in scores over health-related quality of life were noticed to be significantly higher among app users than control subjects in the "My Diabetes Coach" study [45]; thus, providing evidence that such tools are effective in managing not only the clinical symptom improvement but also in emotional function and life satisfaction improvement.

In addition to clinical outcomes and user satisfaction, these mHealth interventions brought positive behavioral changes. Many studies reported that participants became more physically active, followed exercise routines more consistently, and had healthier dietary habits (11/26, 42%). For example, in a diabetes study [68], systolic BP, body fat, and BMI decreased significantly ($P<.001$) among the intervention group, thus indicating the effectiveness of mobile apps in bringing about healthier lifestyle changes.

Finally, 1 study [59] examined the cost-effectiveness of mHealth interventions. It noted that these tools are effective and economically viable, presenting affordable solutions for managing chronic diseases, especially in resource-constrained settings. For example, the chronic pain therapy study [59] established that app-based mobile treatments for pain management were effective and cost-effective, incorporating them into existing treatment plans.

Outcomes of the Formative Studies

The following sections present the findings of the formative studies (n=42) [20,23,25,26,30,32,35,36,38-40,43,44,46-48,51,53,56-58,61,62,70-88]. The most frequent outcomes are patient engagement and satisfaction, but also clinical improvements and self-management behavior are the focus of many studies. There are some usability issues and challenges reported in the studies (Table 5).

Table . Summary of the outcomes and challenges mentioned in the formative studies (n=42).

Outcomes (favorable, no effect, and challenges)	Explanation	Frequency
Favorable outcomes		
Patient engagement and satisfaction (eg, ease of use and positive feedback) [30,32,38-40,44,46,51,58,61,70-78,88]	Most of the studies reported a high level of patient engagement and satisfaction, naming ease of use and personalized feedback as major advantages. Many studies have shown that user-friendly digital health tools improved adherence to health regimens and helped patients manage their diseases more effectively.	22
Clinical improvements (eg, pain reduction, improved blood pressure, and LDL-C[a] levels) [25,32,39,40,43,71,79,80]	These studies reported various clinical improvements, including pain reduction, blood pressure control, improvement in LDL-C level, and overall disease management. Key results included highlighting digital health tools' potential to enhance self-management, support patients to achieve clinical goals, and improve overall health.	8
Self-management behavior (eg, adherence to medication and lifestyle changes) [32,38-40,43,44,46,51,53,56,61,70-73,75,78,81,82]	Overall, digital health tools across these studies helped improve self-management behaviors in medication adherence, lifestyle changes, and engagement in physical activity. The most significant improvements were seen in the management of chronic diseases like COPD[b], diabetes, hypertension, and heart failure.	19
Remote monitoring and resource usage (eg, reduction in hospital visits and health care costs) [25,32,43,51,71,83]	These studies highlight how remote monitoring systems can avert hospital admissions and reduce health costs by helping patients manage their conditions at home and, in real-time, provide the clinician with timely interventions. These findings were seen in heart failure, asthma, COPD management, hypertension, and ankylosing spondylitis, showing how digital health tools could improve clinical outcomes and optimize the use of resources within health care systems.	6
No effect		
No effect on disease management (eg, no improvement in symptom control or disease management) [30,38,48,70,79,84]	Some studies did not find any significant improvement in disease management, especially when the digital health tools did not sufficiently help engage patients or when the patients had barriers such as a lack of interest or limited use of digital health tools. In some cases, the personalization or customization of the tools was insufficient, leading to low effectiveness in managing the conditions.	6
Challenges encountered		
Usability issues (eg, technical issues and user disengagement) [32,35,38,39,61,72,88]	The most common usability issues reported across studies included technical problems, such as device inaccuracies, data syncing issues, and interface complexity. User disengagement was also another common challenge in many instances due to a lack of motivation, the tediousness of the process, and issues relating to poor integration into existing healthcare workflows.	7

Outcomes (favorable, no effect, and challenges)	Explanation	Frequency
Challenges in provider integration (eg, issues with workflows and data sharing) [32,38,43,44,61,71,77,78,81,83]	Provider integration challenges were highly reported in many studies, especially on integrating digital health tools with clinical workflows and the sharing of data between patients and providers. In most cases, the difficulty in adopting digital health tools in routine clinical care was cited as a barrier to clinical decision-making, with issues such as data synchronization and interoperability, assuring that providers can use the data collected remotely efficiently.	10
Privacy and data security concerns (eg, data interoperability and concerns about privacy) [32,38,43,44,61,71,76,81,85]	Privacy and data security issues were consistently identified in the reviews, particularly regarding transmission, storage, and interoperability with other healthcare information systems. There were issues of patient consent, data sharing, and following regulatory policies such as the Health Insurance Portability and Accountability Act (HIPAA). A concern was raised about protecting sensitive health information from unauthorized access.	9
Lack of participation (eg, limited use due to lack of time, motivation, or technical issues) [32,38,43,44,46,51,56,61,70-75,78,81,82]	Several of these studies repeatedly mentioned problems of non-participation for which technical issues, such as malfunction of a device and connectivity problems, together with a lack of motivation, were major reasons for dropouts and inconsistent use of digital health tools. Time constraints were also a significant factor in disengagement, as the patients struggled to integrate such tools into daily life. Personalization and support appeared very pivotal for long-term engagement.	17

[a]LDL-C: low-density lipoprotein cholesterol.

[b]COPD: chronic obstructive pulmonary disease.

One of the significant trends observed across the formative studies is the potential of digital health tools to improve the management of chronic conditions such as COPD, hypertension, diabetes, and heart failure. A significant percentage of studies (17/42, 40%) reported favorable clinical and self-management outcomes, including improved BP control, weight management, and self-efficacy. For example, HbA1c was reduced by approximately 0.79% [57], and patients suffering from COPD on the Wellinks mHealth platform showed improved symptoms and quality of life [71]. Similarly, 1 study [36] demonstrated that digital management tools for hypertension reduced systolic BP/diastolic BP by 14/5 mm Hg. These results suggest that digital tools can offer tangible improvements in managing chronic diseases, particularly when integrated with traditional care methods.

The second trend to emerge from these studies is the centrality of user-centered design in the overall success of digital health interventions. A substantial proportion of studies (13/42, 31%) indicated that designing the digital tool for patient preferences and needs enhances engagement and satisfaction. For example, the iMHere 2.0 system, which offers personalized app modules to support various self-management tasks, was praised for its customizability and ability to keep patients engaged [61]. Similarly, the Wellinks mHealth platform for COPD was well-received due to its ease of use and support in daily disease management [71]. The above findings point out the importance of developing digital tools that are not only functional but can also be tailored according to the needs of the patients in improving usability and increasing engagement.

Whereas the initial engagement and clinical outcomes from the studies were generally good, the longer-term health outcomes tend to be more mixed. The main challenges with the long-term maintenance of digital interventions were mentioned in 12% (5/42) of the selected studies, where initially engaged patients stopped using the tools due to various barriers, such as motivational issues, technical problems, or difficulties in maintaining regular use. For example, some patients in the studies of COPD management dropped off after an initial burst of engagement due to difficulties in integrating the technology into their daily routines [71]. This points to the need for continuous engagement strategies and more user-friendly designs to maintain patient involvement over the long term [26].

The studies also indicate an increased awareness that, in treating chronic conditions of a complex nature, it is more often than not challenging to rely on one-size-fits-all approaches. Some studies with generalized tools showed positive outcomes; others (5/42, 12%) indicated that tools must be customized to meet individual patient needs. For instance, a digital health tool for diabetes showed promising results in improving medication adherence but struggled with user engagement in the long term, particularly among patients who required more personalized support [39]. Many studies emphasized the need for adaptive technologies that can adjust to the changing needs of patients

and those that can integrate seamlessly into existing health care systems.

Despite promising results, the need for further research and development is a constant note in several studies, as shown by 9.52% (4/42). These studies have shown that while the performance of digital tools has a promising side, there are serious gaps in personalization, scalability, and integration into health systems [60,62]. Issues of provider workflow problems [38], data interoperability [38], and assurance about patient privacy concerns were considered the most important to resolve to make them more acceptable.

Discussion

Principal Findings

This review highlights the growing potential of digital platforms in enabling both self-management (patient-facing monitoring, decision support, and behavior change) and provider-led management (remote monitoring dashboards, care coordination, and clinical decision support). The digital platforms were primarily designed for use at home and complement patients' routine health care practices seamlessly, giving major importance to self-monitoring, personalization, and motivational aspects such as rewards. While self-management is valued, our findings suggest that platforms without embedded communication facilities with health care professionals or social support may limit user interaction and effectiveness. The fact that collaborative care and self-management co-occurred in the included studies underscores the value of hybrid interventions that combine patient autonomy with professional oversight.

A significant proportion of the studies reported the usage of co-design or user-centered design approaches as the best practice when developing health interventions. The lack of methodological details and theoretical underpinnings in some studies hinders clear conclusions regarding their effectiveness. While the review found a range of features, there was limited evidence on which features are most effective at facilitating long-term engagement. Most studies were feasibility or pilot studies with brief follow-up periods and small or nonrandomized samples. As such, large-scale assessments of the effectiveness of platforms, particularly for long-term health outcomes, are lacking. Platforms addressing single chronic diseases were more likely to exhibit improved outcomes, whereas those addressing multimorbidity encountered more usability and implementation issues, underscoring the need for more personalized designs to meet complex health needs.

One of the significant barriers to the long-term adoption of digital health solutions was user disengagement. Several longitudinal studies reported high dropout rates due to a lack of motivation or technical issues, highlighting the importance of adaptive and personalized engagement strategies that are unobtrusive and seamlessly integrate digital health platforms into everyday life.

In the following sections, we elaborate on the principal findings and compare and contrast them with relevant literature.

Characteristics of Digital Platforms Supporting the Management of Chronic Diseases

Our findings show that most of the interventions focused on self-management of chronic diseases. We also found that the vast majority of interventions were designed to be used at home by participants, usually in conjunction with their ongoing health care plans. Self-management is critical in managing chronic disease [94,95], and new digital platforms heavily target this aim. According to the literature, chronic disease can be managed well by balancing traditional medical care with self-management [96]. Effective self-management requires optimal communication with health care teams [97]. Although self-management requires support from health care providers and caregivers, this review discovers that not all identified platforms offer such options. Similarly, it is necessary that digital platforms provide options for social support to maintain long-term engagement in self-managing disease [98]. Only a few platforms in our review had options for social support. Our findings suggest that features like self-tracking, customization, and rewards support users' engagement with digital platforms [12]. Although we have identified a range of features in the studied platforms, there is still a lack of evidence in the included studies as to which features are best for supporting long-term engagement. Therefore, more research is needed to investigate which features of digital platforms will best support long-term user engagement and motivation.

The findings from applying the NICE Evidence Standards Framework emphasize the necessity of a multifaceted and integrated approach to delivering effective digital health interventions. The significance of self-management reflects the growing emphasis on patient empowerment, enabling individuals to actively monitor, track, and manage their health. However, self-management alone is not always sufficient, as structured support systems improve engagement and adherence. The frequent co-occurrence of self-management with collaborative care (47 times) highlights the effectiveness of a hybrid model that combines patient autonomy with a form of clinical oversight. This integration ensures that while patients take a leading role in their health management, they are still supported by health care professionals who provide guidance and add to the rigor of the digital intervention. However, as noted earlier, some platforms lack direct communication with health care providers, which could reflect as a limiting factor in their effectiveness as a digital intervention. This comprehensive analysis highlights the interdependencies among digital intervention strategies, advocating for a cohesive, patient-centered approach. Self-management is most effective when complemented by collaborative care, education, and evidence-based tools, ensuring engagement, clinical effectiveness, and long-term sustainability in chronic disease management.

Principles and Frameworks Used in Designing and Developing the Platforms

We have observed that more than half of the studies reported some form of co-design, consultative, or user-centered approach to the development of digital platforms. This is a positive indication that health care interventions are co-developed to

meet the needs of the stakeholders, as suggested in the literature [99,100]. Overall, adopting a co-design approach might have contributed to positive outcomes. However, due to the lack of relevant data in those studies, we could not determine if there was any relationship between the co-design of the digital platforms and their effectiveness in managing chronic disease. A lack of theoretical basis in some of the included studies limits their reliability. Therefore, it is suggested that the future development of digital platforms needs solid theoretical support, and such support may well improve effectiveness and user engagement.

A range of behavior-change techniques was used in the included studies, generally informed by behavior-change models and social cognitive theory [101]. However, future studies should focus on measuring the effects of these design principles, determining the extent to which they contribute to the efficacy and continued use of the platforms. In addition, we identified 5 platforms that used gamification techniques to enhance user experience and embed learning principles. This suggests that platform designers considered the notion that gamification improves health behaviors, as reported in the literature [102].

Effectiveness and Efficacy of the Platforms

Most of the included studies were pilot in nature, focusing either on design, development, usability, uptake, or clinical utility. In many of these studies, not all features of the platforms were tested for effectiveness. One key reason for not testing all the features of a digital platform during the trial period (for example, study by Doyle et al [44]) was that the development was based on the initial success of the trial, and this was the focus of the majority of the included studies. This suggests that future work should identify those studies where the full potential of digital platforms is evaluated, so that more complete conclusions can be drawn about the platforms' effectiveness.

It is important to note that the majority of studies reporting improved health outcomes focused on a single or localized chronic disease. Conversely, those reporting no effects on health outcomes largely focused on multiple chronic diseases. This suggests that focused digital platforms have a higher likelihood of positive health outcomes than those targeting multiple chronic health issues. Further research is needed to investigate how digital platforms could successfully be designed and evaluated to manage multimorbidity (ie, the presence of 2 or more long-term health conditions). Furthermore, it is worth noting that the testing duration of the digital platforms was generally short in most studies. In some cases, the formal sample size was not calculated since the studies were feasibility or pilot studies. This highlights the importance of meticulously evaluating digital platforms with appropriate sample sizes in future studies to ensure the validity and reliability of the research.

Uptake of the Digital Health Interventions

A few of the included studies, which were conducted over a longer period (ie, 6 months or more), explicitly mentioned dropout rates and the causes of disengagement, including lack of motivation, technical issues, or health problems. For example, a study [86] aimed to understand the effects of long-term (eg, 12 months) smartphone-based self-monitoring in patients with lipid metabolism disorders reported that 43% (43/100) of patients never started using the app due to a lack of time, health problems, lack of motivation, and technical problems. Dropout due to technical issues (eg, poor wireless connection) was also reported in another study, where an mHealth system was developed for managing chronic conditions [61], where 1 patient dropped out of the study after the initial engagement. Another included study [39] mentioned that approximately 80% (400/500) of users used Mobiab (ForaCare Suisse AG) for managing diabetes for less than 1 week. Such a dropout may imply that the daily use requirements of the app were challenging for individuals to maintain. User fatigue is a likely issue where daily data entry or constant interaction with digital tools is required. As digital health interventions often require continuous self-monitoring and engagement—whether through tracking symptoms, inputting data, or responding to feedback—users may experience burnout, leading to disengagement. A high dropout rate (95/162, 58.6%) was mentioned in a study [66] where the effect of an mHealth self-monitoring intervention among black individuals with uncontrolled hypertension was tested for 1 year. While digital health tools show promise in improving short-term health outcomes, dropout rates, user fatigue, and sustainability remain substantial challenges. For digital health interventions to be successful in the long run, continuous engagement strategies and adaptive features must be prioritized to ensure users remain motivated and that interventions can be integrated smoothly into everyday life. However, further investigation is required to understand the dropout and disengagement with digital health interventions.

Digital literacies play a key role in the uptake of digital health interventions [103]. Researchers found that even though patients with low literacy may have access to technology, they may not be able to use it without any help [104]. Some studies were conducted with adults aged between early (20 - 39 y) [27,35,43,61] and middle adulthood (40 - 59 y) [30,34,40,55]. Chronic diseases are more common among older adults, who may have poor digital literacy and difficulty adopting new information technologies. In the included studies, no information was collected regarding the digital literacy levels of the participants. However, some of the included studies had information about income and education level that are linked to the participants' digital skills and health literacy. Low digital literacy may be a barrier to the adoption and engagement of digital platforms. Another limitation of the included studies is that not all studies reported parameters such as participants' skills, experience, or level of education, and these might have contributed to the infrequent use of the platform [28] or withdrawal from the study [42]. Some of the studies excluded patients who were unwilling to participate or could not meet the study requirements. Such nonparticipants may provide insights for the design of more user-centered platforms. Therefore, future studies should account for the varying digital literacies of different cohorts of platform users, as this may impact the overall feasibility of digital health interventions.

Participant confidentiality, including data security, was not widely discussed in the papers included in this review. Only 1 paper provided a detailed overview of the data security measures

of their platform [39]. Researchers have reported that health and fitness apps often violate users' privacy by not following existing guidelines and regulations [105,106]. Therefore, it is of the utmost importance to protect the health data that digital platforms gather. Ensuring the privacy and security of health data may contribute to the long-term uptake of digital health interventions.

Implications and Real-World Adoption

While most studies focus on the efficacy of digital health interventions in pilot settings, large-scale integration within already prevailing health care systems was not reported in any studies. However, some studies have promising steps toward integration within the current health care system. As an example, a mobile app called Mobiab [39] for diabetes management integrated with health care systems was partially successful, particularly where patient data could be automatically transmitted to clinicians through platforms like mobile apps. The study reported that the use of different software was an additional complication for the clinicians, as they already used some commercial software. Also, the researchers did not implement data into hospital information systems due to the lack of communication interface specifications. A hypertension management platform was developed that can be integrated into EHRs to facilitate real-time patient data analysis and effective decision-making [85]. However, the efficacy of the system is yet to be tested. Another application called "electronic patient-reported outcome mobile app and portal system" was developed for people with complex care needs. During the trial, the electronic patient-reported outcome system was not interfaced with other existing technology systems, that is, EMRs or other available platforms, but the system was designed in such a way that interoperability could be a possibility [50].

None of the studies explicitly discusses interoperability, clinician workload, or regulatory constraints, although these could hinder the implementation of digital health solutions in real-world health care settings [107,108]. These barriers are implicit in the included studies and require investigation in future studies. For example, one of our included studies [39] refers to the long-term involvement of clinicians in managing diabetes but does not discuss the potential impact it would have on clinician workload. Clinicians would be required to review the data generated by the app, resulting in an increased workload if the app is not providing actionable insights or is not well-integrated into clinical workflows. Research shows that the lack of seamless integration of digital health platforms into health care systems poses obstacles to broader adoption and implementation [109,110] and may hinder sustainability and scalability. In addition, regulatory restrictions on patient data protection (eg, General Data Protection Regulation and Health Insurance Portability and Accountability Act) are not discussed in the included studies, but these would be necessary to make the platform health care regulation compliant and legally viable for real-world adoption and implementation. Such concerns were echoed in a recent study, where authors highlighted the need for transparent data government policies to be implemented in order to meet regulatory requirements and address security concerns [107].

Many of the papers included in this review suggested that further research into the interventions was needed. Reasons for this included limited sample sizes, limited participant uptake, technical issues, and the need for further personalization of the platforms. Furthermore, while short-term health outcomes of the digital interventions were largely positive, long-term outcomes remain generally unknown. Thus, it was frequently suggested that strategies to maintain long-term use were required, along with further analysis of platform use. Additionally, it was noted in some instances that future studies should account for the varying digital literacies of different cohorts of users, as this may impact the overall feasibility of digital health interventions. Finally, where studies focused on a single or localized chronic health issue, it was broadly concluded that platform design could easily be replicated to address other health issues. However, interventions that sought to address multiple chronic diseases had a higher incidence of technical issues or problems with usability and feasibility. This could be because those studies have methodological challenges, such as higher sample sizes and complex sampling frames to measure the required outcomes. However, such challenges were not reported in those studies, and this suggests that further research is needed to enable digital health interventions to effectively address multiple chronic health conditions.

Limitations

There are several limitations to our study. Our objective was to explore the broader perspective of chronic diseases rather than focusing on individual chronic diseases. Therefore, the search strategy used was general terms related to chronic disease management rather than specific conditions such as diabetes, hypertension, cardiovascular disease, or COPD. While this approach allowed us to draw on a wide range of interventions, it may have inadvertently omitted condition-specific intervention studies using narrowly defined keywords. This could have affected the completeness of the review in 2 ways. First, we might have missed highly specialized interventions tailored to the unique management needs of particular chronic diseases. Second, excluding disease-specific search terms could have led to underrepresenting certain populations or technologies in specific disease domains. Nonetheless, studies in this review constitute an important and representative sample of the current digital intervention landscape for chronic disease management. These studies encompass a diverse range of technologies, user groups, and interventions to identify trends, design issues, and gaps in digital interventions for chronic disease management. We acknowledge that a more targeted search strategy, perhaps in a future scoping or systematic review, would offer more information about condition-specific digital health innovations and their impact.

Our search strategy did not include the exact keyword "self-management" or corresponding controlled-vocabulary terms. Although the broader "management" concept and citation chasing were used to capture patient-led self-management studies indirectly, records that exclusively use "self-management" terminology may have been missed. A future update should incorporate a self-management term cluster (eg, "self-management," "self care," and "self-monitor") and mapped controlled vocabulary to improve sensitivity.

In this review, we only included articles that were published in English. For the included studies, we have primarily reported the results qualitatively. Where available, we reported the frequency of outcomes but were unable to capture the effect size due to the variability of the studies. Furthermore, not all studies measured the impact of digital platforms; several of them instead measured the usability and acceptability of those platforms. Therefore, we could not compare all the outcomes, and in some cases, the outcomes were inconclusive due to the preliminary nature of the studies. Finally, the digital platforms included in the studies were designed for diverse users with varying degrees of digital literacy. However, we could not analyze how the digital literacies of the participants contributed to outcomes because such data were not reported in the studies.

Conclusions

This study provides a comprehensive overview of digital platforms for managing chronic diseases, delineating features for self-management versus provider-led management. Overall, the vast majority of papers in this review concluded that digital health interventions can be beneficial in managing chronic health issues. They also indicated that the adoption of such methods in combination with regular clinical care has the potential to improve health outcomes, support self-management, and support communication between patients and health care providers. However, challenges remain in long-term engagement, overcoming technological barriers, and integrating these tools into existing workflows in health care. The effectiveness and acceptance of digital health interventions vary based on patient characteristics, such as age, health literacy, and the capacity for intervention tailoring. Success will be contingent on interventions that can fulfill specific patient needs through user-centered, tailored engagement while being effortless to use and integrated seamlessly within the health care ecosystem. These tools, therefore, require further research for their full development so that they are adaptable, scalable, and meet the diverse needs of patients with chronic conditions. More research is needed to further develop these tools for wider acceptance and improving engagement.

Funding

This work was supported jointly by Peter MacCallum Cancer CentreCenter, Swinburne University of Technology, and the Digital Health CRC Limited (“DHCRC”). DHCRC is funded under the Commonwealth’s Cooperative Research Centres (CRC) Program.

Authors' Contributions

Conceptualization: AT, AAM, SB
Formal analysis: AAM, CL, SJ, MS, ST, ARMF
Funding acquisition: AT, PPJ
Investigation: AAM, CL, SJ
Methodology: AAM, AT, NW, SB
Validation: AAM, SJ, CL
Visualization: AAM, CL, SJ
Writing – original draft: AAM, CL, SJ, PPJ
Writing – review and editing: All authors
All authors reviewed the final manuscript and approved the submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database search outcome.
[DOCX File, 20 KB - mhealth_v14i1e63742_app1.docx]

Multimedia Appendix 2
Chronic diseases reported in the included studies (n=83).
[PNG File, 77 KB - mhealth_v14i1e63742_app2.png]

Multimedia Appendix 3
Digital intervention strategies, features of the digital platforms, and co-occurrence matrix of the digital intervention strategies.
[DOCX File, 50 KB - mhealth_v14i1e63742_app3.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 86 KB - mhealth_v14i1e63742_app4.docx]

References

1. Chronic disease. US Centers for Disease Control and Prevention. 2024. URL: https://www.cdc.gov/chronic-disease/about/index.html#:~:text=Chronic%20diseases%20are%20defined%20broadly,of%20daily%20living%20or%20both [accessed 2024-11-11]
2. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2023-11-13]
3. The top 10 causes of death. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2023-11-14]
4. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133. [Medline: 25642490]
5. Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. Prev Med Rep 2018 Dec;12:284-293. [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
6. Murray E, Hekler EB, Andersson G, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
7. Samal L, Fu HN, Camara DS, Wang J, Bierman AS, Dorr DA. Health information technology to improve care for people with multiple chronic conditions. Health Serv Res 2021 Oct;56 Suppl 1(Suppl 1):1006-1036. [doi: 10.1111/1475-6773.13860] [Medline: 34363220]
8. Anekwe TD, Rahkovsky I. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2018 Dec;108(S6):S430-S436. [doi: 10.2105/AJPH.2014.302041r]
9. Sheng Y, Doyle J, Bond R, Jaiswal R, Gavin S, Dinsmore J. Home-based digital health technologies for older adults to self-manage multiple chronic conditions: a data-informed analysis of user engagement from a longitudinal trial. Digit Health 2022;8:20552076221125957. [doi: 10.1177/20552076221125957] [Medline: 36171962]
10. What is digital health technology and what can it do for me? National Institute of Health and Care Research. 2022. URL: https://evidence.nihr.ac.uk/collection/what-is-digital-health-technology/ [accessed 2025-12-14]
11. Bashi N, Fatehi F, Mosadeghi-Nik M, Askari MS, Karunanithi M. Digital health interventions for chronic diseases: a scoping review of evaluation frameworks. BMJ Health Care Inform 2020 Mar;27(1):e100066. [doi: 10.1136/bmjhci-2019-100066] [Medline: 32156751]
12. Nittas V, Zecca C, Kamm CP, Kuhle J, Chan A, von Wyl V. Digital health for chronic disease management: an exploratory method to investigating technology adoption potential. PLoS ONE 2023;18(4):e0284477. [doi: 10.1371/journal.pone.0284477] [Medline: 37053272]
13. Digital health platform handbook: building a digital information infrastructure (infostructure) for health. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/9789240013728 [accessed 2025-12-14]
14. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
15. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143. [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
16. Tricco AC, Lillie E, Zarin W, 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]
17. Covidence. URL: https://www.covidence.org/ [accessed 2025-12-05]
18. Gagnon MP, 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]
19. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
20. Genberg EM, Viitanen HT, Mäkelä MJ, Kautiainen HJ, Kauppi PM. Impact of a digital web-based asthma platform, a real-life study. BMC Pulm Med 2023 May 12;23(1):165. [doi: 10.1186/s12890-023-02467-8] [Medline: 37173716]
21. Goodman KS, Locke E. Health impact of a mobile-delivered diabetes intervention to control blood pressure in older adults. AJPM Focus 2024 Aug;3(4):100244. [doi: 10.1016/j.focus.2024.100244] [Medline: 39034932]
22. Abel WM, Efird JT, Crane PB, Ferdinand KC, Foy CG, DeHaven MJ. Use of coaching and technology to improve blood pressure control in Black women with hypertension: pilot randomized controlled trial study. J of Clinical Hypertension 2023 Jan;25(1):95-105. [doi: 10.1111/jch.14617]
23. Chrysohoou C, Tsamadias V, Kariori M, et al. Evaluation of the usability of the digital platform navigator KardioUp for the journey of patients with chronic heart failure. Hellenic J Cardiol 2024;75:32-40. [doi: 10.1016/j.hjc.2023.06.001] [Medline: 37295668]
24. Xu W, Huang X, Lin Q, et al. Application of Alfalfa app in the management of oral anticoagulation in patients with atrial fibrillation: a multicenter randomized controlled trial. BMC Med Inform Decis Mak 2024 Oct 9;24(1):294. [doi: 10.1186/s12911-024-02701-1] [Medline: 39385171]
25. Reid D, Mehta J, Anis K, Mehta S. Impact of remote patient monitoring platform on patients with moderate to severe persistent asthma: observational study. JMIR Form Res 2023 Dec 28;7:e51065. [doi: 10.2196/51065] [Medline: 38153783]

26. Alharbey R, Chatterjee S. An mHealth assistive system "MyLung" to empower patients with chronic obstructive pulmonary disease: design science research. JMIR Form Res 2019 Mar 19;3(1):e12489. [doi: 10.2196/12489] [Medline: 30888329]
27. Kryger MA, Crytzer TM, Fairman A, et al. The effect of the interactive mobile health and rehabilitation system on health and psychosocial outcomes in spinal cord injury: randomized controlled trial. J Med Internet Res 2019 Aug 28;21(8):e14305. [doi: 10.2196/14305] [Medline: 31464189]
28. Puig J, Echeverría P, Lluch T, et al. A specific mobile health application for older HIV-infected patients: usability and patient's satisfaction. Telemed J E Health 2021 Apr;27(4):432-440. [doi: 10.1089/tmj.2020.0098] [Medline: 32667858]
29. Chhabra HS, Sharma S, Verma S. Smartphone app in self-management of chronic low back pain: a randomized controlled trial. Eur Spine J 2018 Nov;27(11):2862-2874. [doi: 10.1007/s00586-018-5788-5] [Medline: 30324496]
30. Evans J, Papadopoulos A, Silvers CT, et al. Remote health monitoring for older adults and those with heart failure: adherence and system usability. Telemed J E Health 2016 Jun;22(6):480-488. [doi: 10.1089/tmj.2015.0140] [Medline: 26540369]
31. Dorsch MP, Farris KB, Rowell BE, Hummel SL, Koelling TM. The effects of the ManageHF4Life mobile app on patients with chronic heart failure: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 7;9(12):e26185. [doi: 10.2196/26185] [Medline: 34878990]
32. Guo X, Gu X, Jiang J, et al. A hospital-community-family-based telehealth program for patients with chronic heart failure: single-arm, prospective feasibility study. JMIR Mhealth Uhealth 2019 Dec 13;7(12):e13229. [doi: 10.2196/13229] [Medline: 31833835]
33. Velardo C, Shah SA, Gibson O, et al. Digital health system for personalised COPD long-term management. BMC Med Inform Decis Mak 2017 Feb 20;17(1):19. [doi: 10.1186/s12911-017-0414-8] [Medline: 28219430]
34. Oh SW, Kim KK, Kim SS, Park SK, Park S. Effect of an integrative mobile health intervention in patients with hypertension and diabetes: crossover study. JMIR Mhealth Uhealth 2022 Jan 11;10(1):e27192. [doi: 10.2196/27192] [Medline: 35014961]
35. Setiawan IMA, Zhou L, Alfikri Z, et al. An adaptive mobile health system to support self-management for persons with chronic conditions and disabilities: usability and feasibility studies. JMIR Form Res 2019 Apr 25;3(2):e12982. [doi: 10.2196/12982] [Medline: 31021324]
36. 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]
37. Schnall R, Cho H, Mangone A, Pichon A, Jia H. Mobile health technology for improving symptom management in low income persons living with HIV. AIDS Behav 2018 Oct;22(10):3373-3383. [doi: 10.1007/s10461-017-2014-0] [Medline: 29299790]
38. Gill A, Khan AI, Hans PK, Kuluski K, Cott C. The electronic patient reported outcome tool: testing usability and feasibility of a mobile app and portal to support care for patients with complex chronic disease and disability in primary care settings. JMIR Mhealth Uhealth 2016 Jun 2;4(2):e58. [doi: 10.2196/mhealth.5331] [Medline: 27256035]
39. Burda V, Mráz M, Schneider J, Novák D. Managing diabetes using Mobiab: long-term case study of the impact of a mobile app on self-management. JMIR Diabetes 2022 Apr 20;7(2):e36675. [doi: 10.2196/36675] [Medline: 35442201]
40. Bailey JF, Agarwal V, Zheng P, et al. Digital care for chronic musculoskeletal pain: 10,000 participant longitudinal cohort study. J Med Internet Res 2020 May 11;22(5):e18250. [doi: 10.2196/18250] [Medline: 32208358]
41. Zhang P, Tao X, Ma Y, et al. Improving the management of type 2 diabetes in China using a multifaceted digital health intervention in primary health care: the SMARTDiabetes cluster randomised controlled trial. Lancet Reg Health West Pac 2024 Aug;49:101130. [doi: 10.1016/j.lanwpc.2024.101130] [Medline: 39056088]
42. Poppe L, De Bourdeaudhuij I, Verloigne M, et al. Efficacy of a self-regulation-based electronic and mobile health intervention targeting an active lifestyle in adults having type 2 diabetes and in adults aged 50 years or older: two randomized controlled trials. J Med Internet Res 2019 Aug 2;21(8):e13363. [doi: 10.2196/13363] [Medline: 31376274]
43. Ji X, Wang Y, Ma Y, et al. Improvement of disease management and cost effectiveness in Chinese patients with ankylosing spondylitis using a smart-phone management system: a prospective cohort study. Biomed Res Int 2019;2019:2171475. [doi: 10.1155/2019/2171475] [Medline: 30931322]
44. Doyle J, Murphy E, Gavin S, et al. A digital platform to support self-management of multiple chronic conditions (ProACT): findings in relation to engagement during a one-year proof-of-concept trial. J Med Internet Res 2021 Dec 15;23(12):e22672. [doi: 10.2196/22672] [Medline: 34914612]
45. Gong E, Baptista S, Russell A, et al. My Diabetes Coach, a mobile app-based interactive conversational agent to support type 2 diabetes self-management: randomized effectiveness-implementation trial. J Med Internet Res 2020 Nov 5;22(11):e20322. [doi: 10.2196/20322] [Medline: 33151154]
46. Babington-Ashaye A, de Moerloose P, Diop S, Geissbuhler A. Design, development and usability of an educational AI chatbot for people with haemophilia in Senegal. Haemophilia 2023 Jul;29(4):1063-1073. [doi: 10.1111/hae.14815] [Medline: 37347648]
47. Barbaric A, Munteanu C, Ross H, Cafazzo JA. Design of a patient voice app experience for heart failure management: usability study. JMIR Form Res 2022 Dec 6;6(12):e41628. [doi: 10.2196/41628] [Medline: 36472895]
48. Breckner A, Litke N, Göbl L, et al. Effects and processes of an mHealth intervention for the management of chronic diseases: prospective observational study. JMIR Form Res 2022 Aug 25;6(8):e34786. [doi: 10.2196/34786] [Medline: 36006666]

49. Lear SA, Norena M, Banner D, et al. Assessment of an interactive digital health-based self-management program to reduce hospitalizations among patients with multiple chronic diseases: a randomized clinical trial. JAMA Netw Open 2021 Dec 1;4(12):e2140591. [doi: 10.1001/jamanetworkopen.2021.40591] [Medline: 34962560]
50. Steele Gray C, Chau E, Tahsin F, et al. Assessing the implementation and effectiveness of the electronic patient-reported outcome tool for older adults with complex care needs: mixed methods study. J Med Internet Res 2021 Dec 2;23(12):e29071. [doi: 10.2196/29071] [Medline: 34860675]
51. Rossetto F, Borgnis F, Isernia S, et al. System integrated digital empowering and telerehabilitation to promote patient activation and well-being in chronic disabilities: a usability and acceptability study. Front Public Health 2023;11:1154481. [doi: 10.3389/fpubh.2023.1154481] [Medline: 37250091]
52. Hartch CE, Dietrich MS, Lancaster BJ, Stolldorf DP, Mulvaney SA. Effects of a medication adherence app among medically underserved adults with chronic illness: a randomized controlled trial. J Behav Med 2024 Jun;47(3):389-404. [doi: 10.1007/s10865-023-00446-2] [Medline: 38127174]
53. Salari R, R Niakan Kalhori S, GhaziSaeedi M, Jeddi M, Nazari M, Fatehi F. Mobile-based and cloud-based system for self-management of people with type 2 diabetes: development and usability evaluation. J Med Internet Res 2021 Jun 2;23(6):e18167. [doi: 10.2196/18167] [Medline: 34076579]
54. Pfaeffli Dale L, Whittaker R, Jiang Y, Stewart R, Rolleston A, Maddison R. Text message and internet support for coronary heart disease self-management: results from the Text4Heart randomized controlled trial. J Med Internet Res 2015 Oct 21;17(10):e237. [doi: 10.2196/jmir.4944] [Medline: 26490012]
55. Sittig S, Wang J, Iyengar S, Myneni S, Franklin A. Incorporating behavioral trigger messages into a mobile health app for chronic disease management: randomized clinical feasibility trial in diabetes. JMIR Mhealth Uhealth 2020 Mar 16;8(3):e15927. [doi: 10.2196/15927] [Medline: 32175908]
56. Nabovati E, Rangraz Jeddi F, Tabatabaeizadeh SM, Hamidi R, Sharif R. Design, development, and usability evaluation of a smartphone-based application for nutrition management in patients with type II diabetes. J Diabetes Metab Disord 2023;22(1):315-323. [doi: 10.1007/s40200-022-01140-x]
57. NOVIANI L, DIANTINI A, SUBARNAS A. Implementation of mobile application for diabetes self-management on self efficacy and clinical outcome in Indonesia. International Journal of Pharmaceutical Research (09752366) 2020 [FREE Full text]
58. Duan H, Wang Z, Ji Y, et al. 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 Feb 25;8(2):e14466. [doi: 10.2196/14466] [Medline: 32130161]
59. Morcillo-Muñoz Y, Sánchez-Guarnido AJ, Calzón-Fernández S, Baena-Parejo I. Multimodal chronic pain therapy for adults via smartphone: randomized controlled clinical trial. J Med Internet Res 2022 May 11;24(5):e36114. [doi: 10.2196/36114] [Medline: 35373776]
60. Park SK, Bang CH, Lee SH. Evaluating the effect of a smartphone app-based self-management program for people with COPD: a randomized controlled trial. Appl Nurs Res 2020 Apr;52:151231. [doi: 10.1016/j.apnr.2020.151231] [Medline: 31955942]
61. Parmanto B, Pramana G, Yu DX, Fairman AD, Dicianno BE, McCue MP. iMHere: a novel mHealth system for supporting self-care in management of complex and chronic conditions. JMIR Mhealth Uhealth 2013 Jul 11;1(2):e10. [doi: 10.2196/mhealth.2391] [Medline: 25100682]
62. Brandl LC, Liebram C, Schramm W, Pobiruchin M. A German smartphone-based self-management tool for psoriasis: community-driven development and evaluation of quality-of-life effects. JMIR Form Res 2022 Jul 7;6(7):e32593. [doi: 10.2196/32593] [Medline: 35797109]
63. Li C, Huang J, Wu H, et al. Management of rheumatoid arthritis with a digital health application: a multicenter, pragmatic randomized clinical trial. JAMA Netw Open 2023 Apr 3;6(4):e238343. [doi: 10.1001/jamanetworkopen.2023.8343] [Medline: 37058302]
64. Goulding EH, Dopke CA, Rossom R, Jonathan G, Mohr D, Kwasny MJ. Effects of a smartphone-based self-management intervention for individuals with bipolar disorder on relapse, symptom burden, and quality of life: a randomized clinical trial. JAMA Psychiatry 2023 Feb 1;80(2):109-118. [doi: 10.1001/jamapsychiatry.2022.4304] [Medline: 36542401]
65. Zhang W, Zhu X, Que X, Zhang X. The effects of shared decision-making informed dietary intervention based on digital health technology in older adults with type 2 diabetes mellitus: a randomized controlled trial. Digit Health 2024;10:20552076241272514. [doi: 10.1177/20552076241272514] [Medline: 39403713]
66. Buis LR, Kim J, Sen A, et al. The effect of an mHealth self-monitoring intervention (MI-BP) on blood pressure among black individuals with uncontrolled hypertension: randomized controlled trial. JMIR Mhealth Uhealth 2024 Jun 28;12:e57863. [doi: 10.2196/57863] [Medline: 38941601]
67. Tabernero C, Gutiérrez-Domingo T, Steca P, et al. Effectiveness of mindfulness and positive strengthening mHealth interventions for the promotion of subjective emotional wellbeing and management of self-efficacy for chronic cardiac diseases. J Pers Med 2022 Nov 25;12(12):1953. [doi: 10.3390/jpm12121953] [Medline: 36556174]

68. Patnaik L, Panigrahi SK, Sahoo AK, Mishra D, Muduli AK, Beura S. Effectiveness of mobile application for promotion of physical activity among newly diagnosed patients of type II diabetes - a randomized controlled trial. Int J Prev Med 2022;13(1):54. [doi: 10.4103/ijpvm.IJPVM_92_20] [Medline: 35706879]
69. Jia W, Zhang P, Zhu D, et al. Evaluation of an mHealth-enabled hierarchical diabetes management intervention in primary care in China (ROADMAP): a cluster randomized trial. PLoS Med 2021 Sep;18(9):e1003754. [doi: 10.1371/journal.pmed.1003754] [Medline: 34547030]
70. Burka D, Gupta R, Moran AE, et al. Keep it simple: designing a user-centred digital information system to support chronic disease management in low/middle-income countries. BMJ Health Care Inform 2023 Jan;30(1):e100641. [doi: 10.1136/bmjhci-2022-100641] [Medline: 36639189]
71. Gelbman BD, Reed CR. An integrated, multimodal, digital health solution for chronic obstructive pulmonary disease: prospective observational pilot study. JMIR Form Res 2022 Mar 17;6(3):e34758. [doi: 10.2196/34758] [Medline: 35142291]
72. Rusch A, Carley I, Badola P, et al. Digital mental health interventions for chronic serious mental illness: findings from a qualitative study on usability and scale-up of the life goals app for bipolar disorder. Front Digit Health 2022;4:1033618. [doi: 10.3389/fdgth.2022.1033618] [Medline: 36479190]
73. Hietbrink EAG, Middelweerd A, van Empelen P, et al. A digital lifestyle coach (E-Supporter 1.0) to support people with type 2 diabetes: participatory development study. JMIR Hum Factors 2023 Jan 12;10(1):e40017. [doi: 10.2196/40017] [Medline: 36633898]
74. O'Neill K, Gormley C, Kelly MG, et al. Service development project to pilot a digital technology innovation for video direct observation of therapy in adult patients with asthma. BMJ Open Qual 2024 Jul 15;13(3):e002626. [doi: 10.1136/bmjoq-2023-002626] [Medline: 39009461]
75. Salim H, Cheong AT, Sharif-Ghazali S, et al. A self-management app to improve asthma control in adults with limited health literacy: a mixed-method feasibility study. BMC Med Inform Decis Mak 2023 Sep 27;23(1):194. [doi: 10.1186/s12911-023-02300-6] [Medline: 37759184]
76. Barbaric A, Munteanu C, Ross H, Cafazzo JA. A voice app design for heart failure self-management: proof-of-concept implementation study. JMIR Form Res 2022 Dec 21;6(12):e40021. [doi: 10.2196/40021] [Medline: 36542435]
77. Díaz-Mohedo E, Carrillo-León AL, Calvache-Mateo A, Ptak M, Romero-Franco N, Carlos-Fernández J. App-Mohedo®: a mobile app for the management of chronic pelvic pain. A design and development study. Int J Med Inform 2024 Jun;186:105410. [doi: 10.1016/j.ijmedinf.2024.105410] [Medline: 38507980]
78. Kim DY, Kwon H, Nam KW, Lee Y, Kwon HM, Chung YS. Remote management of poststroke patients with a smartphone-based management system integrated in clinical care: prospective, nonrandomized, interventional study. J Med Internet Res 2020 Feb 27;22(2):e15377. [doi: 10.2196/15377] [Medline: 32130140]
79. Xie X, Wang X, Li A, et al. A study of the effectiveness of mobile health application in a self-management intervention for kidney transplant patients. Iran J Kidney Dis 2023 Sep;17(5):263-270. [doi: 10.52547/ijkd.7693] [Medline: 37838936]
80. Holtz BE, Mitchell KM, Holmstrom AJ, et al. The effect of an mHealth intervention for adolescents with type 1 diabetes and their parents. J Telemed Telecare 2024 Aug;30(7):1155-1162. [doi: 10.1177/1357633X221125835] [Medline: 36177538]
81. Xing F, Guo Y, Xia N, et al. Mobile app-assisted family physician program for improving blood pressure outcome in hypertensive patients. BMC Prim Care 2023 Jan 10;24(1):8. [doi: 10.1186/s12875-023-01965-2] [Medline: 36627556]
82. Miki T, Yamada J, Ishida S, Sakui D, Kanai M, Hagiwara Y. Exploring the feasibility and initial impact of an mHealth-based disease management program for chronic ischemic heart disease: formative study. JMIR Form Res 2024 Aug 22;8:e56380. [doi: 10.2196/56380] [Medline: 39173150]
83. Naranjo-Rojas A, Perula-de Torres L, Cruz-Mosquera FE, Molina-Recio G. Usability of a mobile application for the clinical follow-up of patients with chronic obstructive pulmonary disease and home oxygen therapy. Int J Med Inform 2023 Jul;175:105089. [doi: 10.1016/j.ijmedinf.2023.105089] [Medline: 37172506]
84. Lalloo C, Hundert A, Harris L, et al. Capturing daily disease experiences of adolescents with chronic pain: mHealth-mediated symptom tracking. JMIR Mhealth Uhealth 2019 Jan 17;7(1):e11838. [doi: 10.2196/11838] [Medline: 30664472]
85. Funes Hernandez M, Babakhanian M, Chen TP, et al. Design and implementation of an electronic health record-integrated hypertension management application. J Am Heart Assoc 2024 Jan 16;13(2):e030884. [doi: 10.1161/JAHA.123.030884] [Medline: 38226516]
86. Steinert A, Eicher C, Haesner M, Steinhagen-Thiessen E. Effects of a long-term smartphone-based self-monitoring intervention in patients with lipid metabolism disorders. Assist Technol 2020;32(2):109-116. [doi: 10.1080/10400435.2018.1493710] [Medline: 29944463]
87. Fee C, Fuller J, Guss CE, et al. A digital platform to support HIV case management for youth and young adults: mixed methods feasibility study. JMIR Form Res ;6(11):e39357. [doi: 10.2196/39357]
88. Cormican O, Dowling M. Evaluating a digital self-management tool for people living with multiple myeloma. Cancer Nursing Practice 2020;20(1):36-42 [FREE Full text]
89. Digital Literacy. 2024. URL: https://literacy.ala.org/digital-literacy [accessed 2024-11-05]
90. Health literacy. World Health Organization. 2024. URL: https://www.who.int/teams/health-promotion/enhanced-wellbeing/ninth-global-conference/health-literacy [accessed 2024-11-05]

91. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2025-12-14]
92. National framework for the integrated prevention and management of chronic disease in Ireland 2020–2025. : Health Service Executive; 2020 URL: https://www.hse.ie/eng/about/who/cspd/icp/chronic-disease/documents/national-framework-integrated-care.pdf [accessed 2025-12-14]
93. Van de Velde D, De Zutter F, Satink T, et al. Delineating the concept of self-management in chronic conditions: a concept analysis. BMJ Open 2019 Jul 16;9(7):e027775. [doi: 10.1136/bmjopen-2018-027775] [Medline: 31315862]
94. Modi AC, Pai AL, Hommel KA, et al. Pediatric self-management: a framework for research, practice, and policy. Pediatrics 2012 Feb;129(2):e473-e485. [doi: 10.1542/peds.2011-1635] [Medline: 22218838]
95. Grande SW, Longacre MR, Palmblad K, et al. Empowering young people living with juvenile idiopathic arthritis to better communicate with families and care teams: content analysis of semistructured interviews. JMIR Mhealth Uhealth 2019 Feb 22;7(2):e10401. [doi: 10.2196/10401] [Medline: 30794202]
96. Lorig K. Self-management of chronic illness: a model for the future. Generations: Journal of the American Society on Aging 1993;17(3):11-14 [FREE Full text]
97. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
98. Bellin MH, Dosa N, Zabel TA, Aparicio E, Dicianno BE, Osteen P. Self-management, satisfaction with family functioning, and the course of psychological symptoms in emerging adults with spina bifida. J Pediatr Psychol 2013;38(1):50-62. [doi: 10.1093/jpepsy/jss095] [Medline: 22976508]
99. Lupton D. The digitally engaged patient: self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Aug;11(3):256-270. [doi: 10.1057/sth.2013.10]
100. 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. [doi: 10.2196/jmir.4055] [Medline: 25639757]
101. Schunk DH, Ellen LU. Social cognitive theory and motivation. In: The Oxford Handbook of Human Motivation: Oxford University Press; 2012. [doi: 10.1093/oxfordhb/9780195399820.013.0002]
102. Al-Rayes S, Al Yaqoub FA, Alfayez A, et al. Gaming elements, applications, and challenges of gamification in healthcare. Informatics in Medicine Unlocked 2022;31:100974. [doi: 10.1016/j.imu.2022.100974]
103. Kemp E, Trigg J, Beatty L, et al. Health literacy, digital health literacy and the implementation of digital health technologies in cancer care: the need for a strategic approach. Health Prom J of Aust 2021 Feb;32(S1):104-114 [FREE Full text] [doi: 10.1002/hpja.387]
104. Vollbrecht H, Arora V, Otero S, Carey K, Meltzer D, Press VG. Evaluating the need to address digital literacy among hospitalized patients: cross-sectional observational study. J Med Internet Res 2020 Jun 4;22(6):e17519. [doi: 10.2196/17519] [Medline: 32496196]
105. Papageorgiou A, Strigkos M, Politou E, Alepis E, Solanas A, Patsakis C. Security and privacy analysis of mobile health applications: the alarming atate of practice. IEEE Access 2018;6:9390-9403. [doi: 10.1109/ACCESS.2018.2799522]
106. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the far side of mobile health: information security and privacy of mobile health apps on iOS and Android. JMIR Mhealth Uhealth 2015 Jan 19;3(1):e8. [doi: 10.2196/mhealth.3672] [Medline: 25599627]
107. Alzghaibi H. Healthcare practitioners' perceptions of mHealth application barriers: challenges to adoption and strategies for enhancing digital health integration. Healthcare (Basel) 2025 Feb 25;13(5):494. [doi: 10.3390/healthcare13050494] [Medline: 40077056]
108. Leigh S, Ashall-Payne L, Andrews T. Barriers and facilitators to the adoption of mobile health among health care professionals from the United Kingdom: discrete choice experiment. JMIR Mhealth Uhealth 2020 Jul 6;8(7):e17704. [doi: 10.2196/17704] [Medline: 32628118]
109. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223. [doi: 10.2196/11223] [Medline: 30990458]
110. Giebel GD, Abels C, Plescher F, et al. Problems and barriers related to the use of mHealth apps from the perspective of patients: focus group and interview study. J Med Internet Res 2024 Apr 23;26:e49982. [doi: 10.2196/49982] [Medline: 38652508]

Abbreviations

AI: artificial intelligence
BP: blood pressure
COPD: chronic obstructive pulmonary disease
EHR: electronic health record
EMR: electronic medical record

HbA1c: hemoglobin A1c
mHealth: mobile health
NICE: National Institute for Health and Care Excellence
PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by L Buis; submitted 06.Jul.2024; peer-reviewed by A Dima, B Seljelid, N Nøhr; revised version received 06.Sep.2025; accepted 01.Oct.2025; published 14.Jan.2026.

Please cite as:
Al Mahmud A, Joachim S, Jayaraman PP, Learmonth C, Tyagi S, Forkan ARM, Shuakat M, Wickramasinghe N, Wheeler J, Best S, Trainer A
Digital Health Interventions to Support Chronic Disease Management: Systematic Scoping Review
JMIR Mhealth Uhealth 2026;14:e63742
URL: https://mhealth.jmir.org/2026/1/e63742
doi:10.2196/63742

Mobile Apps to Improve Health Parameters in Healthy Adults: Systematic Review

Gaia Leuzzi[1,2], PhD; Mirko Job[1], PhD; Carola Cosentino[3], PhD; Riccardo Testa[4], MSc; Aldo Scafoglieri[2], PhD; Marco Testa[1], PhD

[1]Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Campus of Savona, Savona, Italy

[2]Department of Physical Education and Rehabilitation, Experimental Anatomy Research Group (EXAN), Vrije Universiteit of Brussel (VUB), Brussels, Belgium

[3]Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Genova, Italy

[4]Department of Electrical, Electronics, Telecommunication Engineering and Naval Architecture, University of Genoa, Genova, Italy

Corresponding Author:
Marco Testa, PhD
Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, Campus of Savona, Savona, Italy

Abstract

Background: Recently, mobile health and mobile apps have been proposed as a potential tool to improve different outcomes (eg, daily steps, blood glucose) in both people with and without chronic conditions. In particular, healthy people could benefit from these tools by improving health variables and for prevention. Previous evidence investigated different types of health interventions adopting apps in various settings and populations, but evidence of their effectiveness is still unclear.

Objective: The aim was to assess the effectiveness of mobile apps in improving health variables (eg, daily steps, maximal aerobic capacity) in healthy adults, involving an intervention regarding physical activity, diet, or their combination thereof. Evidence would suggest if apps could be effectively adopted in health interventions aiming toward prevention.

Methods: A systematic review was performed using Medline via PubMed, Cochrane Library—CENTRAL, and Embase. Only randomized controlled trials comparing the same intervention provided with and without a mobile app or a treatment and a mobile app compared with the treatment only were included in this systematic review. The Risk of Bias tool 2.0 was used to assess the risk of bias, and the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) was adopted for rating the certainty of evidence.

Results: Considering studies up to June 2025, only 2 studies were included in the review of mobile apps for physical activity, and none were included for mobile apps for diet and none for mobile apps for physical activity and diet combined. The quality of evidence of the 2 studies included was low due to a high risk of bias, several missing data, and deviation from the original interventions, suggesting a scarce rigor in the methodology adopted. Therefore, mobile apps' effectiveness in improving diet, physical activity, or their combination cannot be assessed.

Conclusions: Despite the widespread use of mobile apps for health and the large number of relative publications, the results of this systematic review did not allow us to ascertain the effectiveness of mobile apps for health, but they provided fundamental insights for future research. Hence, it is not possible to state if apps for health might be used as supporting tools for health interventions aiming toward prevention and health improvements in healthy people. There is an urgent need to develop stronger evidence of apps' effectiveness in addressing different populations and types of interventions for different health domains.

Trial Registration: PROSPERO CRD42023485803; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023485803

(JMIR Mhealth Uhealth 2026;14:e66881) doi:10.2196/66881

KEYWORDS

active aging; diet; exercise; healthy ageing; health apps; mobile apps; physical activity

Introduction

Promoting active aging [1] requires new strategies to reach different populations in a feasible and effective way to help people improve their health status [2,3]. For this reason, the use of technology interventions is getting paramount attention to help people improve their health variables [4], identified as both clinical (eg, blood pressure, weight) and nonclinical outcomes (eg, daily activity or sleep) [5,6]. The attention to the use of technology-supported health interventions is due to the ease of

the use of mobile devices, their portability, their quality-price ratio [7], and the quantity of information they can provide with good data storage and live data analysis [8].

The World Health Organization describes this type of app as mobile health app (mHealth), defining it as "medical and public health practices supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices" [9]. Wireless devices can be fitness trackers, smartwatches, and smartphones, which makes it easy to collect several types of data (ie, number of daily steps, macronutrients, sleep, and stress level) from different health-related spheres, such as physical activity and diet [10,11], automatically (ie, using a wearable device [12]), through user's action or both. In particular, to do so, smartphones allow for downloading different types of apps, including health ones. These apps can be focused on just 1 health-related sphere [13,14] or a combination thereof [15].

Due to their easy use and versatility, people with and without diseases can benefit from mHealth apps to prevent or treat different conditions [16,17]. In line with that, different studies have been carried out in the last few years to test the effectiveness of apps for health on people with [18] and without diseases [19]. Unfortunately, the quality of the studies is low, and the results are often controversial [20], thus not allowing for definitive results on this topic. Moreover, to understand the effect of apps for health, one should compare the effectiveness of an intervention provided with and without the adoption of an app, while most of these studies adopt apps as baseline treatment with different interventions as adjunctive therapy, therefore testing the effectiveness of the adjunctive therapy rather than the app one [21,22]. Furthermore, mobile apps were chosen as mHealth to be investigated, as they are one of the most adopted technologies worldwide in health contexts [23], and those targeting fitness, nutrition, and healthy living are widely diffused [24].

Hence, the main purpose of this systematic review is to analyze the effectiveness of mobile apps in improving healthy adults' (ie, >18 years old) health variables, analyzing only randomized controlled trials (RCTs) that compare the same intervention with and without this technology in physical activity, diet, and a combination thereof.

Methods

The protocol of this systematic review was created and submitted to PROSPERO [25] (CRD42023485803). Furthermore, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Checklist 1) [26] guidelines and PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension; Checklist 2) [27] were followed to report this review.

Deviations from the Protocol

The protocol initially restricted inclusion to English-language studies. During screening, it became clear that this would underrepresent available evidence and introduce potential language bias. We therefore expanded the criteria to include studies in other languages when reliable translation or accurate data extraction was possible. This deviation was made to enhance the review's completeness and international representativeness, while maintaining all other methodological criteria.

Study Objective

The main objective of this systematic review was to analyze the effectiveness of mobile apps for physical activity and diet to improve healthy adults' health variables. The research question was as follows: are mobile apps effective in improving health-related variables in healthy adults? To address this topic, three different options were investigated: (1) mobile apps for physical activity, (2) mobile apps for diet, and (3) mobile apps for physical activity and diet. The main outcomes listed hereafter correspond to each of the abovementioned points: (1) physical activity variables (eg, daily steps, moderate-to-vigorous physical activity), (2) diet variables (eg, weight, BMI), and (3) physical activity and diet variables as mentioned above.

Eligibility Criteria

For this systematic review, studies were considered eligible if they were published RCTs. No limitations on publication time were set, and RCTs published online until June 3, 2025, were included. Systematic reviews, reviews, meta-analyses, single-case studies, case series, observational studies, books, documents, guidelines, reports, and conference abstracts were excluded. Gray literature, systematic reviews, and meta-analyses were consulted, but not considered eligible, to find useful studies.

We included all those studies that involved healthy adult participants (>18 years old), with no cognitive impairments, musculoskeletal or neuromotor diseases, chronic conditions (eg, diabetes, hypertension), obese (ie, BMI≥30), or pregnant women. Interventions were considered eligible if a mobile app was used as an intervention to improve variables related to physical activity, diet, or their combination or only as a supportive technology to a specific intervention. No limitations on time or sessions of interventions were set. Mobile apps, including automatic or self-reported data collection and the use of a wearable device, were taken into account. Conversely, we excluded interventions with non-standalone apps, using high-cost sensors, exergames, nonmobile monitoring systems, robotics systems, or just clinician's telemedicine. Moreover, studies that did not precisely describe the health conditions of their population were excluded.

Search Strategy

Three scientific databases were sought for the study research: Medline via PubMed, Cochrane Library—CENTRAL, and Embase. They were chosen as they are reported as mandatory by the *Cochrane Handbook for Systematic Reviews of Interventions* [28].

The literature search was performed on the databases up to June 3, 2025, and the results were later merged into a single file to be subsequently uploaded onto Covidence [29], where the automatic duplicate detection was conducted. Specific search strings were created for the 3 databases, mixing Boolean

operators (ie, AND, OR), MeSH terms, and keywords. The research strategy is reported below.

- Mobile apps and physical activity and RCT: ((mHealth) OR (m-health) OR ("mobile health") OR ("mobile application") OR ("mobile app") OR ("smartphone application") OR ("smartphone app") OR (apps) OR (smartphone) OR ("Mobile Applications"[Mesh]) OR ("Smartphone"[Mesh])) AND ((fitness) OR ("physical exercise") OR ("physical fitness") OR ("fitness behavior") OR ("Physical Fitness"[Mesh]) OR ("Exercise"[Mesh]) OR (pedometer) OR (steps) OR (exercise) OR ("training exercise") OR ("heart rate variability") OR ("Heart rate") OR ("Heart Rate"[Mesh])) AND ((single blind) OR (double blind) OR (trial) OR (random*) OR (randomized) OR (randomized controlled))
- Mobile apps and diet and RCT: ((mHealth) OR (m-health) OR ("mobile health") OR ("mobile application") OR ("mobile app") OR ("smartphone application") OR ("smartphone app") OR (apps) OR (smartphone) OR ("Mobile Applications"[Mesh]) OR ("Smartphone"[Mesh])) AND ((diet) OR ("calorie counter") OR ("calorie counting") OR (calorie) OR ("calorie intake") OR (diet) OR ("Diet"[Mesh]) OR (dieting) OR ("weight loss") OR ("weight loss"[Mesh]) OR ("Weight Reduction Programs"[Mesh])) AND ((single blind) OR (double blind) OR (trial) OR (random*) OR (randomized) OR (randomized controlled))

Selection Process

Regarding the first research question, 2 researchers (GL and MJ) manually and independently screened titles and abstracts of the retrieved papers and evaluated them against the inclusion criteria. At the same time, for the second research question, 2 researchers (GL and RT) followed the same procedure. For the third research question, studies were identified among the papers selected by the above-described screening. The eligibility of the studies was then agreed upon through a consensus meeting between the 2 authors of each review and, in case of disagreement, a third researcher (CC) was consulted to reach a final decision. Afterward, the full texts of the selected papers were further screened against the inclusion criteria following the same process.

Data Collection

Two researchers (GL and MJ and GL and RT) proceeded blindly and independently to extract specific data from each study such as authors, year of publication, country, intervention setting, study design, total number of participants, number of participants for each experimental group, mean age of the participants and standard deviation (if available), number of female and male participants, type and timing of intervention sessions for both experimental groups, number and timing of follow-ups, outcomes, key conclusions, and eventually even a researcher's comment on each study. Moreover, all data available in each study were extracted and reported, such as mean, median, IQR, SD, number of follow-ups, and data registered at each follow-up. In case of missing data, authors were contacted.

Data Items

The most relevant characteristics of the selected studies are summarized in Table 1. According to our research questions, the outcomes of this systematic review are grouped as follows: (1) physical activity outcomes, (2) diet outcomes, and (3) physical activity and diet outcomes. No limits were identified for the reporting of any outcome. In case of missing data, authors were contacted.

Table . Study characteristics.

Study characteristics	Zongpa et al [30] (2020)	Muntaner-Mas et al [31] (2021)
Total number of participants	47	66
App name	Take a Walk	Vidahora
Type of intervention	App+diet indications	App
Type of control	No app+diet indications	No app
Primary outcome	VO_2 max[a] + HRV[b] + FBG[c] + adherence	Weigh + waist and hip circumference + 20-m shuttle run test + handgrip + standing long jump test + 4 × 10 m shuttle run test + sit and reach + IFIS[d]
Follow-ups	Week 4	Week 9
Intervention sessions and duration	6 reminders/day, 1 for every working hour	Free

[a]VO_2 max: maximal aerobic capacity.

[b]HRV: heart rate variability.

[c]FBG: fasting blood glucose.

[d]IFIS: International Fitness Scale.

Risk of Bias Assessment

The risk of bias (RoB) assessment was performed, independently and blindly, by 2 researchers for each study (GL and MJ or GL and RT, respectively) following the Revised Cochrane Risk of Bias tool 2.0 (RoB 2.0) [32] for RCTs or the Rob 2 CRT for cluster-randomized controlled trials [33]. This tool aims at assessing the RoB specifically for 5 domains: "Risk of bias arising from the randomization process," "Risk of bias due to

deviations from the intended interventions," "Risk of bias due to missing outcome data," "Risk of bias in measurement of outcome," and "Risk of bias in selection of the reported results." Consequently, an overall RoB for the study is provided. Domains and studies can be classified at low, moderate, or high RoB. The tool also allows one to indicate "no information" as an answer for each item of every domain and, in this case, it would often be considered at high RoB. A third researcher (CC) was contacted in case of disagreement to reach a consensus.

Statistical Analysis

Data from each study were extracted and reported, and a descriptive statistic was performed. For intergroup comparisons, the mean, SD, and/or mean differences for pre- and posttreatment conditions were reported. Additionally, the 1- or 2-tailed *t* tests for normally distributed data and the Mann-Whitney *U* test for nonnormally distributed data were also reported if performed in the studies. All statistical analyses were performed using the Jamovi statistical software [34].

Quality of Evidence

To perform the quality of evidence assessment, the GRADE (Grading of Recommendations, Assessment, Development and Evaluation) [35] approach was used via the GRADEpro GTD tool. This tool helps assess both the certainty of evidence and the strength of recommendations. The evaluation process took into account 5 different domains: risk of bias, imprecision (eg, sample size, confidence intervals), inconsistency (eg, heterogeneity), indirectness (eg, eligibility criteria against actual studies included), and publication bias (eg, bias in results publication).

Results

Study Selection

The literary search process for the first review identified a total of 13,444 studies. Duplicate removal eliminated 3436 studies, leaving a total of 10,008 studies to screen. After applying the inclusion and exclusion criteria to titles and abstracts, 51 studies were left [19,30,31,36-82]. Full-text studies were read independently by 2 researchers (GL and MJ), and in due course, another 49 papers were excluded [19,36-82], resulting in the final inclusion of 2 studies for further analysis [30,31]. The complete research process is graphically displayed in Figure 1, and the reasons for exclusions are reported in Multimedia Appendix 1. Multimedia Appendix 2 reports the complete research processes of the other 2 research questions.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for mobile apps for physical activity.

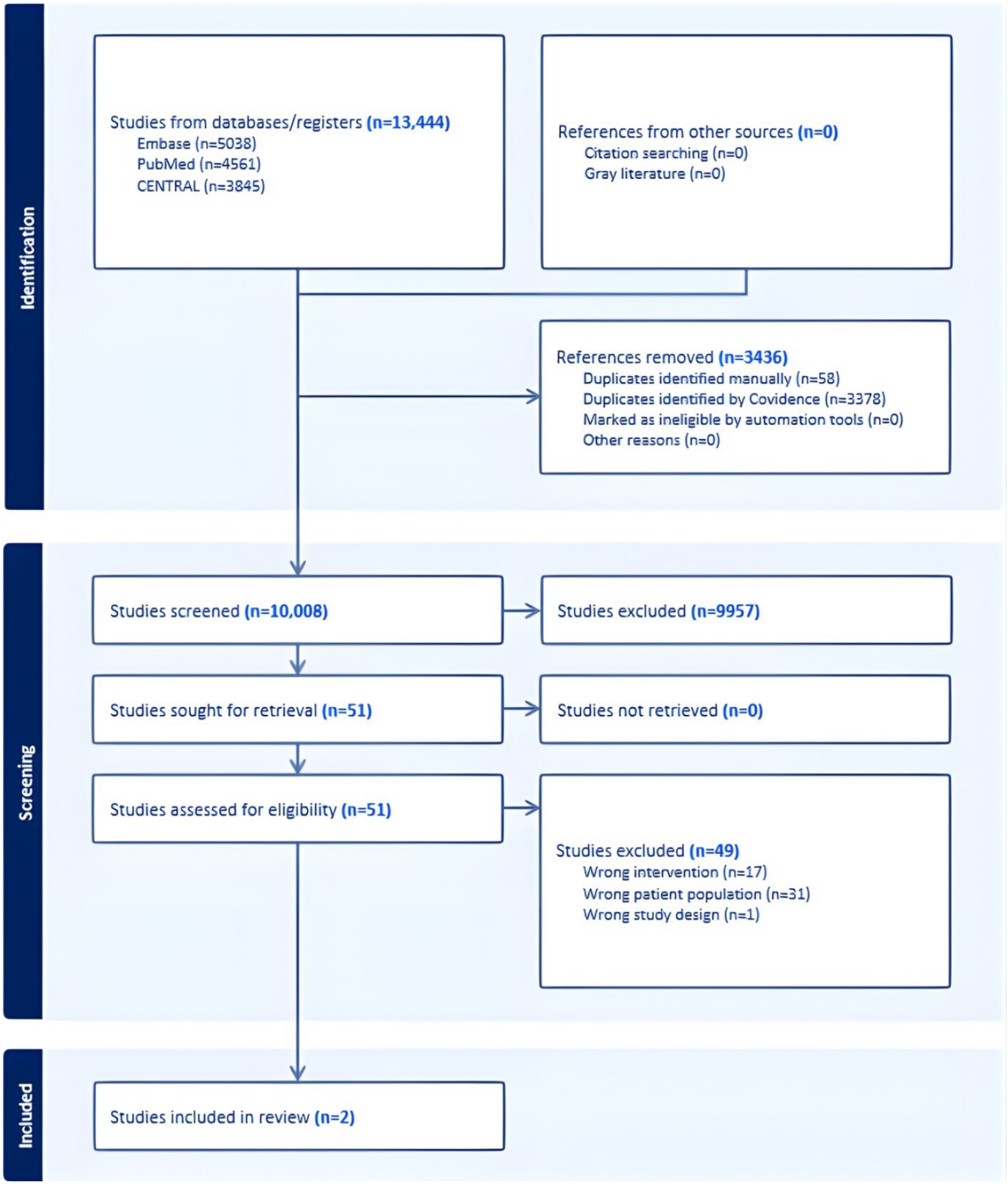
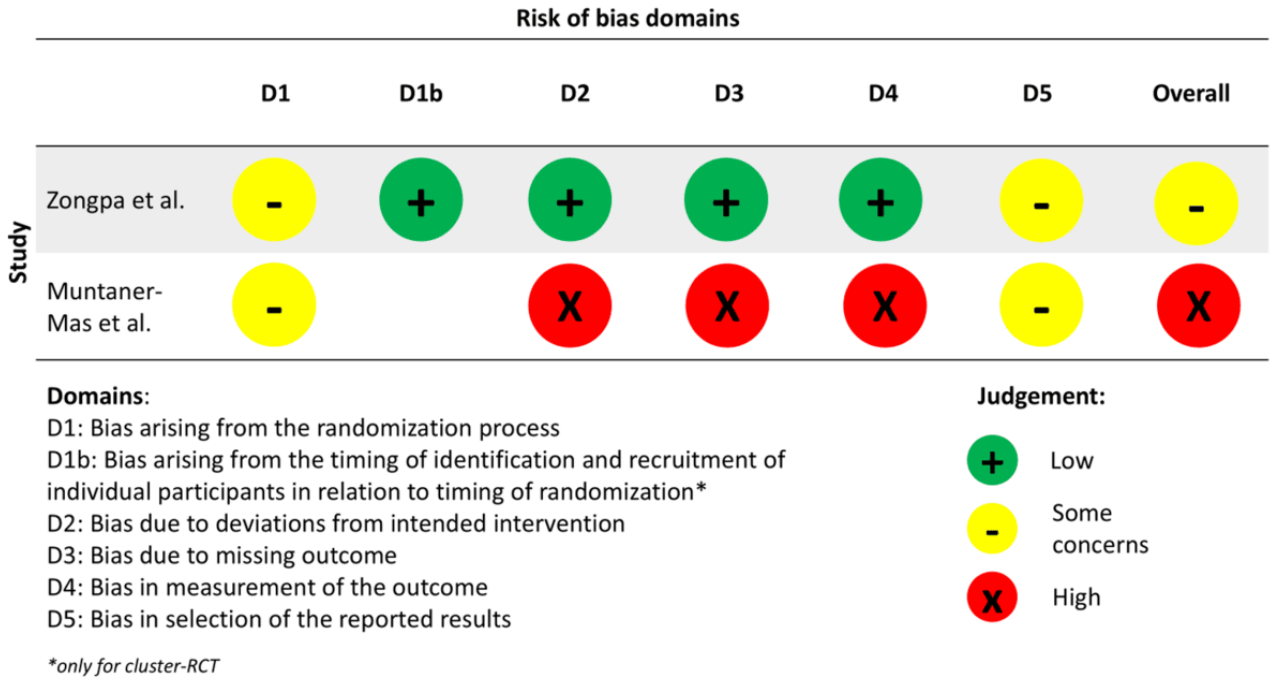


Figure 2. Risk of bias (RoB) assessment for randomized controlled trial (RCT) and cluster-RCT [30,31].



Regarding the second and third questions (ie, mobile app for diet, and physical activity and diet), unfortunately, no studies were included. The selection processes are shown in Multimedia Appendix 2.

Study Characteristics

Among the 2 studies included in the final analysis, the first was a cluster RCT published in 2020 [30], and the second one was an RCT published in 2021 [31]. Both had only 1 intervention group and 1 control group [30,31]. The time of the intervention ranged from 4 to 9 weeks. The countries of study development were India [30] and Spain [31]. The studies' characteristics are indicated in Table 1. The first study [30] involved 47 participants, and the intervention consisted of giving a smartphone app (Take a Walk) to remind them to perform a few minutes of walking at regular intervals. The intervention group received "walk breaks" reminders each hour of work for a total of 6 reminders a day. The intervention lasted 4 weeks. Both intervention and control groups received an indication to follow a standard diet of 2300 Kcal/day. Physical activity–related variables were evaluated using different tests or variables, such as the fasting blood glucose, VO_2 max (maximal aerobic capacity), heart-rate variability, and adherence to walking breaks.

In the second study [31], involving 66 participants, the intervention group was provided with an app (Vidahora) to be used for 9 weeks. Participants were free to use the app whenever they wanted and were only advised to perform at least 10 minutes of physical activity 3 times a day. The second study assessed physical activity–related variables using several performance tests, anthropometric tests, and a questionnaire. Anthropometric tests were weight, hips, and waist circumferences. Performance tests were 20-m shuttle run test, handgrip test, standing long jump test, 4×10 m shuttle run test, and sit and reach test. The questionnaire used was the International Fitness Scale [83].

The characteristics of both studies are reported in Table 1, while the studies' outcomes are reported in Table 2.

Table . Physical activity-related outcomes.

Authors (app used) and test	Baseline		Post-intervention	
	Intervention group	Control group	Intervention group	Control group
Zongpa et al [30] (Take a Walk)				
VO_2 max[a] (mL/kg/min), median (IQR)	45.3 (39.0-52.3)	36.0 (36.0-41.2)	47.6 (39.6-55.9)	37.5 (35-40.3)
HRV[b]				
Time domain, median (IQR)				
SDNN[c] interval (ms)	52.7 (51.4-53.4)	49.1 (44.8-52.1)	54.5 (52.5-60.2)	49.4 (48.1-50.6)
RMSSD[d] interval (ms)	52.4 (48.2-54.2)	54.2 (52.4-55.8)	58.1 (57.4-58.6)	55.1 (53.2-55.7)
NN50[e] (beats)	129.4 (129.1-129.7)	129.7 (128.9-129.2)	130.3 (129.6-132.1)	126.2 (123.8-127.9)
pNN50[f] (%)	37.4 (37.3-37.6)	37.2 (37.0-37.2)	39.0 (38.2-39.2)	35.3 (34.2-37.1)
Frequency domain, median (IQR)				
VLF[g] (ms^2/Hz)	103.00 (101.0-105.0)	103.0 (90.0-103.0)	107.0 (102.0-112.0)	90.1 (84.7-103.1)
LF[h] (ms^2/Hz)	981.0 (972.0-988.2)	984.3 (983.6-985.7)	986.5 (978.0-996.0)	976.0 (962.0-984.0)
HF[i] (ms^2/Hz)	970.0 (958.0-984.0)	981.5 (965.0-984.0)	986.0 (973.0-994.5)	975.0 (962.0-982.0)
LF/HF[j] (%)	1.02 (0.98 - 1.04)	1.0 (0.96 - 1.04)	1.0 (0.9 - 1.0)	1.0 (0.98 - 1.04)
Nonlinear index, median (IQR)				
SD1[k] (ms)	37.2 (36.1-38.3)	22.1 (22.0-25.1)	41.1 (38.2-44.1)	35.8 (34.5-37.7)
SD2[k] (ms)	55.0 (51.8-59.7)	33.5 (30.7-33.5)	88.6 (72.0-88.6)	59.4 (57.6-61.1)
SD1/SD2 (%)	1.5 (1.2-1.8)	1.5 (1.2-1.5)	2.1 (1.9-2.2)	1.7 (1.3-2.0)
FBG[l] (mmol/dL), median (IQR)	89.0 (78.0-93.2)	87.0 (81.0-88.0)	83.0 (72.0-86.0)	87.0 (81.0-92.1)
Adherence	NA[m]	NA	NA	NA
Muntaner-Mas et al [31] (Vidahora)				
Weight (kg), mean (SD)	65.1 (12.1)	65.1 (13.7)	65.7 (12.1)	65.5 (13.8)
Waist circumference (cm), mean (SD)	77.1 (9.9)	81.0 (11.8)	76.0 (11.7)	80.2 (11.2)
Hip circumference (cm), mean (SD)	96 (8.8)	96.3 (9.0)	93.3 (9.1)	95.7 (10.2)
20-m shuttle run (laps), mean (SD)	6.6 (3.2)	5.8 (3.4)	7.7 (2.8)	5.7 (3.3)
Handgrip strength (kg), mean (SD)	30.8 (8.0)	28.9 (7.9)	32.1 (9.0)	28.2 (9.1)
Standing broad jump (cm), mean (SD)	155.4 (35.2)	146.0 (31.0)	169.7 (35.9)	150.5 (29.4)
4×10 m shuttle run (sec), mean (SD)	11.4 (1.3)	11.7 (1.4)	11.3 (1.3)	12.0 (1.5)
Sit-and-reach (cm), mean (SD)	19.9 (8.9)	20.2 (8.9)	21.6 (9.7)	21.7 (8.3)
General physical fitness, mean (SD)	3.2 (0.7)	3.3 (0.9)	3.6 (0.6)	3.2 (0.9)
Cardiorespiratory fitness, mean (SD)	2.8 (1.0)	2.8 (1.1)	3.1 (1.0)	2.7 (1.1)

Authors (app used) and test	Baseline		Post-intervention	
	Intervention group	Control group	Intervention group	Control group
Muscular fitness, mean (SD)	3.2 (0.7)	0.1 (1.1)	3.4 (0.7)	3.2 (1.0)
Speed-agility, mean (SD)	3.3 (0.7)	3.3 (1.0)	3.5 (0.7)	3.3 (1.0)
Flexibility, mean (SD)	2.8 (0.9)	2.8 (1.2)	3.1 (1.2)	2.9 (1.1)

[a]VO_2 max: maximal aerobic capacity.

[b]HRV: heart rate variability.

[c]SDNN interval: standard deviation of NN intervals.

[d]RMSSD: root mean square of successive RR interval differences.

[e]NN50: successive RR intervals that differ by more than 50 ms.

[f]pNN50: percentage of successive RR intervals that differ by more than 50 ms.

[g]VLF: very low frequency of power.

[h]LF: absolute power of the low-frequency band (0.04-0.15 Hz).

[i]HF: absolute power of the high-frequency band (0.15-0.4 Hz).

[j]LF/HF: ratio of LF-HF power.

[k]SD1 and SD2: Poincaré plots perpendicular to line of identity.

[l]FBG: fasting blood glucose.

[m]NA: not available but requested to authors.

App's Characteristics

Information about the Take a Walk app was limited; it was described as a simple Java–based Android app that allowed participants to set personalized reminders for walking, including customizable times and data. Participants were only required to manually set when to receive the reminders to walk.

The Vidahora app was made of 4 different sections: the first section dedicated to a quiz about healthy habits, the second section dedicated to the challenges for improving different physical activities' components via suggested video exercises (eg, strength, aerobic exercise, yoga), the third section hosted an artificial intelligence–assisted chatbot that could ask the participant about progress in a friendly way, and the last section was for setting the user data (eg, username, personal data). Badges for achievements were also present, as well as individual and community challenges with daily or weekly aims. Participants in the intervention groups were invited to use the app as they wished, with the only suggestion of recording at least 3 sessions of a minimum of 10 minutes per week of physical education.

Risk of Bias in Studies

The RoB assessment for the included RCT studies is graphically reported in Figure 2 using the Robvis tool [84]. The first study presents some concerns in the overall RoB, due to the randomization process since it is not clear how the experimenters performed it. Moreover, the study reports the registration of a protocol with a registration number, but in the mentioned database, it is not possible to find the protocol. In this case, it cannot be excluded that the results were not analyzed by a prespecified analysis plan that was finalized before unblinded outcome data were available for analysis. Indeed, the authors reported having several missing data but without any reasonable explanation.

The same 2 domains (D1 and D5) influence the overall RoB of the second study as well. The second study has the overall RoB indicated as "high risk." Domains 2, 3, and 4 are at high RoB, and those are influencing the overall RoB. The corresponding authors of the studies were contacted, but no answer was ever received.

GRADE Assessment

The assessment of the quality of evidence adopting the GRADE approach could not be performed due to the high heterogeneity of the outcomes considered in the studies included. First, the 2 studies did not consider the same outcome. Moreover, even involving the same sphere of interest (ie, physical activity) did not consider the same outcomes. Specifically, the first study [30] mainly considered physiological outcomes (ie, VO_2 max, blood glucose), while the second study [31] investigated performance outcomes (eg, 20-m shuttle run test, handgrip). Therefore, it is not possible to assess the quality of this evidence.

Discussion

Principal Findings

The included studies were overall characterized by a high RoB due to many missing values, high dropout, small sample sizes, and poor data reporting. In particular, the randomization processes were evaluated with the RoB 2.0 tool with "some concerns" as this tool requires this scoring if the paper does not describe the randomization procedures adopted in detail but just mention their adoption. Additionally, for 1 study, a protocol was not available. Moreover, the RoB of the second study was influenced by the decision of not including in the data analysis the participants who had missing data in 1 of the evaluations or were outliers. Thus, it might not be excluded that the results were influenced.

Furthermore, their study designs were significantly different, and it was not possible to make a direct comparison between their results. Our findings highlight the urgent need for standardized outcome measures to enable the generation of stronger, comparable evidence in this field. For each health domain examined, a validated set of standardized outcomes should be developed, allowing for their consistent use across diverse study designs and settings, including RCTs. Such standardization would facilitate more accurate assessments of mobile app effectiveness on health outcomes in healthy adults. Due to the heterogeneity of outcome measures among the included studies, it remains difficult to draw definitive conclusions regarding the efficacy of health-related mobile apps in this population. Therefore, future studies should aim to include larger sample sizes to enhance statistical power and improve the reliability of findings. Additionally, greater participant numbers may also help mitigate issues related to dropout and incomplete data during interventions. Moreover, intervention times should also be standardized to be able to compare results from different studies. Another problem that emerges from our results is the lack of standardized apps or guidelines to develop them for different health domains, and this might be seen from the different outcomes considered in each study included in our review. A standardized version of the health app could allow for having a set of common health data across different apps, with the possibility of adding other health variables specific to each app based on its characteristics and aims. In this way, studies could compare the use of different health apps for the same domain and consider a minimum set of common health variables.

Other systematic reviews were carried out on mHealth in the last years [85-89] and were also characterized by a very limited number of selected studies with high heterogeneity, and therefore they could not assess mHealth effectiveness. Indeed, many of the studies they included presented mixed results of the delivered interventions and the way of delivering them [64,87,90,91]. Hence, those studies did not assess the effectiveness of the same intervention delivered with and without mHealth, as we conversely did in our work. Considering the available literature, it is fundamental to emphasize the need to evaluate the effectiveness of specific mHealth interventions. However, the intrinsic variability in the designs and the scarce quality of the currently available studies do not allow us to state if mobile apps can be effective to improve health variables.

To overcome this problem, we decided to include in our work the studies that compared the same intervention provided with and without the mobile app, and additionally also the studies considering the same treatment provided via mobile app against the treatment. By applying these severe criteria, many studies were excluded from this systematic review for improper control, leaving only 2 studies to analyze. Consistent with our findings, other works reported the need for more studies with clearer designs to test the effectiveness of mHealth technology in different settings [92-95].

Studies Included

Digging into the included studies, Zongpa et al reported that physical activity–related variables improved over a 4-week period, specifically VO_2 max, heart rate variability, and fasting blood glucose. It should be noted that dropouts were 11.32%, and VO_2 max improvements may be questioned because the validity of a submaximal test in healthy people is questionable. Moreover, the VO_2 max improvement reported by this study was only 1.33%, while the minimal clinically important difference for VO_2 max should be higher than 6% [96]. Finally, the results should be taken carefully, as the results included in the analysis considered only an intervention adherence of at least 70%, thus imposing a possible bias in the selection of the results.

In the second study, Muntaner-Mas et al reported that many physical activity–related variables evaluated in the study improved, and the authors decided to split them into 3 categories: fatness indicators, physical fitness components, and self-reported fitness.

Starting from fatness indicators (ie, weight, waist, and hip circumference), no changes were obtained that could be attributable to the app. Physical fitness components (ie, 20-m shuttle run laps, handgrip strength, standing broad jump, 4×10 m shuttle run, and sit-and-reach) were improved, but even though few changes are indicated as statistically significant, the actual improvements are minimal and might not even be clinically relevant [97]. Second, improvements of a few units of centimeters or seconds obtained in 9 weeks and from a healthy and young population could be considered scarce. The category “self-reported physical fitness” explored 5 domains, and all the components were evaluated via Likert-type questions and reported the results obtained from the International Fitness Scale. The results improved for the intervention group and decreased for the control group, but considering the possible bias emerging from the self-evaluation, improvements should be carefully addressed since changes were minimal. Even in this category, the results on mobile app effectiveness could not be considered conclusive.

Despite the study considering the mobile app effective in improving physical activity–related variables, there were many missing data (ie, about 30%), and the sample size is limited. Summarizing, the effectiveness of the Vidahora app cannot be assessed. Although the selected studies highlighted that diet and physical activity levels can be improved by mHealth apps, their weak methodological design raises some concerns about their conclusions.

Limitation

A limitation of this work must be acknowledged: this systematic review included only studies involving healthy participants. Therefore, it is not possible to report anything about people with chronic conditions or pathologies. Additionally, a librarian was not consulted to develop the research strings, as people with expertise in conducting systematic reviews, and in their methodology, were consulted.

Conclusions

Despite the studies we included seeming to support the effectiveness of mobile apps to improve physical activity, diet-related variables, or their combination in healthy adults,

their poor methodological quality as well as the high variability in literature does not allow any definitive conclusion on this topic. Besides, the long-term effects of mobile apps interventions on different outcomes are scarce [92,95], and further research is needed. In addition, the interventions' (eg, activity, diet) effectiveness should be tested a priori and then provided via mHealth. Some urgent needs emerge from the literature analyzed and from this study. Specifically, for future studies, there is a need for high-quality RCT designs with large sample sizes to better assess the possible effects of health apps and the generalizability of results. Moreover, there is a need for clearer and consistent methodology that could provide stronger evidence of effectiveness, more transparent reporting of results that would prevent any bias and would additionally allow for acknowledging what is not working with apps for health and why, and addressing healthy people to test the mobile apps' effectiveness in preventing diseases and improving health conditions. Furthermore, standardized outcomes for each health domain of interest (eg, physical activity, diet) should be adopted, allowing for comparing the results of different studies and populations. Additionally, different studies could include, in the same health domain of interest, the chosen standardized outcomes as well as new ones, to try to expand possible results. Moreover, mHealth should be tested and validated by both patients and users before using them to deliver an intervention. Further research should test mobile apps as a tool supporting preventive approaches for health and well-being in young people and healthy participants as well. Finally, clear guidelines should be created on how to build up different types of mHealth, specifically for mobile apps, to standardize this process among health apps and to further try to ensure better use of this technology in the active aging and well-being fields.

Acknowledgments

We would like to sincerely express our gratitude to Dr Simone Battista for his support in carrying out this work. In all parts of this manuscript, generative artificial intelligence was used solely to ensure clarity and correct syntax. It was not used to create or contribute any insights into the research itself or any of its components.

Funding

No funding was received for this study.

Data Availability

All the relevant details and information are presented in the manuscript. Nonetheless, the data are available on reasonable request from the authors.

Authors' Contributions

GL, CC, MJ, and MT performed study conceptualization. GL and CC performed methodology. GL, RT, and MJ conducted the analysis, investigation, and data extraction. GL wrote the original draft preparation. GL, MJ, CC, RT, AS, and MT reviewed and edited it. All authors read and approved the final manuscript.

Multimedia Appendix 1
Reasons for exclusions (mobile apps for physical activity).
[DOCX File, 19 KB - mhealth_v14i1e66881_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagrams.
[DOCX File, 54 KB - mhealth_v14i1e66881_app2.docx]

Checklist 1
PRISMA checklist.
[DOCX File, 22 KB - mhealth_v14i1e66881_app3.docx]

Checklist 2
PRISMA-S checklist.
[DOCX File, 30 KB - mhealth_v14i1e66881_app4.docx]

References

1. Walker A. A strategy for active ageing. Int Soc Secur Rev 2002 Jan;55(1):121-139. [doi: 10.1111/1468-246X.00118]
2. Cai T, Verze P, Bjerklund Johansen TE. The quality of life definition: where are we going? Uro 2021 Feb;1(1):14-22. [doi: 10.3390/uro1010003]

3. Noble N, Paul C, Turon H, Oldmeadow C. Which modifiable health risk behaviours are related? A systematic review of the clustering of Smoking, Nutrition, Alcohol and Physical activity ('SNAP') health risk factors. Prev Med 2015 Dec;81:16-41. [doi: 10.1016/j.ypmed.2015.07.003] [Medline: 26190368]
4. Lee M, Lee H, Kim Y, et al. Mobile app-based health promotion programs: a systematic review of the literature. Int J Environ Res Public Health 2018 Dec 13;15(12):2838. [doi: 10.3390/ijerph15122838] [Medline: 30551555]
5. Leuzzi G, Recenti F, Giardulli B, Scafoglieri A, Testa M. Exploring digital health: a qualitative study on adults' experiences with health apps and wearables. Int J Qual Stud Health Well-being 2025 Dec 31;20(1). [doi: 10.1080/17482631.2024.2447096]
6. Measures of health status, quality of life and health care. Health Knowledge. URL: https://www.healthknowledge.org.uk/public-health-textbook/research-methods/1c-health-care-evaluation-health-care-assessment/measures-health-status [accessed 2024-03-19]
7. Randazzo V, Ferretti J, Pasero E. Anytime ECG monitoring through the use of a low-cost, user-friendly, wearable device. Sensors (Basel) 2021 Sep 9;21(18):6036. [doi: 10.3390/s21186036] [Medline: 34577247]
8. Al-Rawashdeh M, Keikhosrokiani P, Belaton B, Alawida M, Zwiri A. IoT adoption and application for smart healthcare: a systematic review. Sensors (Basel) 2022 Jul 19;22(14):5377. [doi: 10.3390/s22145377] [Medline: 35891056]
9. Park YT. Emerging new era of mobile health technologies. Healthc Inform Res 2016 Oct;22(4):253-254. [doi: 10.4258/hir.2016.22.4.253] [Medline: 27895955]
10. Thornton L, Gardner LA, Osman B, et al. A multiple health behavior change, self-monitoring mobile app for adolescents: development and usability study of the Health4Life app. JMIR Form Res 2021 Apr 12;5(4):e25513. [doi: 10.2196/25513] [Medline: 33843590]
11. Leuzzi G, Job M, Scafoglieri A, Testa M. Smartphone Apps and Wearables for Health Parameters in Young Adulthood: Cross-Sectional Study. JMIR Hum Factors ;12:e64629-e64629. [doi: 10.2196/64629]
12. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med 2021 Mar 10;4(1):45. [doi: 10.1038/s41746-021-00418-3] [Medline: 33692479]
13. Slazus C, Ebrahim Z, Koen N. Mobile health apps: an assessment of needs, perceptions, usability, and efficacy in changing dietary choices. Nutrition 2022 Sep;101:111690. [doi: 10.1016/j.nut.2022.111690] [Medline: 35660502]
14. Nardi W, Roy A, Dunsiger S, Brewer J. Analyzing the impact of mobile app engagement on mental health outcomes: secondary analysis of the unwinding anxiety program. J Med Internet Res 2022 Aug 15;24(8):e33696. [doi: 10.2196/33696] [Medline: 35969440]
15. Ruf A, Koch ED, Ebner-Priemer U, Knopf M, Reif A, Matura S. Studying microtemporal, within-person processes of diet, physical activity, and related factors using the APPetite-mobile-app: feasibility, usability, and validation study. J Med Internet Res 2021 Jul 5;23(7):e25850. [doi: 10.2196/25850] [Medline: 34342268]
16. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e21759. [doi: 10.2196/21759] [Medline: 33074161]
17. Sarker MHR, Moriyama M, Rashid HU, et al. Chronic kidney disease awareness campaign and mobile health education to improve knowledge, quality of life, and motivation for a healthy lifestyle among patients with chronic kidney disease in Bangladesh: randomized controlled trial. J Med Internet Res 2022 Aug 11;24(8):e37314. [doi: 10.2196/37314] [Medline: 35969429]
18. Bylappa BK, Kamath DY, Josephine IS, et al. Usability and feasibility assessment of a smartphone application (Suhriday) for heart failure self-care remote monitoring in an Indian tertiary health care setting: a pilot mixed-methods study. BMJ Open 2022 Aug 24;12(8):e056962. [doi: 10.1136/bmjopen-2021-056962] [Medline: 36002201]
19. Al-Nawaiseh HK, McIntosh WA, McKyer LJ. An-m-health intervention using smartphone app to improve physical activity in college students: a randomized controlled trial. Int J Environ Res Public Health 2022 Jun 13;19(12):7228. [doi: 10.3390/ijerph19127228] [Medline: 35742477]
20. Edwards D, Williams J, Carrier J, Davies J. Technologies used to facilitate remote rehabilitation of adults with deconditioning, musculoskeletal conditions, stroke, or traumatic brain injury: an umbrella review. JBI Evid Synth 2022 Aug 1;20(8):1927-1968. [doi: 10.11124/JBIES-21-00241] [Medline: 35971198]
21. Ek A, Alexandrou C, Söderström E, et al. Effectiveness of a 3-month mobile phone-based behavior change program on active transportation and physical activity in adults: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jun 8;8(6):e18531. [doi: 10.2196/18531] [Medline: 32510462]
22. ANZCTR - Registration. Australian New Zealand Clinical Trials Registry (ANZCTR). URL: https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377945&showOriginal=true&isReview=true [accessed 2022-10-21]
23. Khan R, Khan S, Almohaimeed HM, Almars AI, Pari B. Utilization, challenges, and training needs of digital health technologies: perspectives from healthcare professionals. Int J Med Inform 2025 May;197:105833. [doi: 10.1016/j.ijmedinf.2025.105833] [Medline: 39954392]
24. Wang Y, Wu T, Chen Z. Active usage of mobile health applications: cross-sectional study. J Med Internet Res 2021 Dec 22;23(12):e25330. [doi: 10.2196/25330] [Medline: 34941545]
25. Schiavo JH. PROSPERO: an international register of systematic review protocols. Med Ref Serv Q 2019;38(2):171-180. [doi: 10.1080/02763869.2019.1588072] [Medline: 31173570]

26. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71. [doi: 10.1136/bmj.n71] [Medline: 33782057]
27. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. Syst Rev 2021 Jan 26;10(1):39. [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
28. Higgins JPT, Thomas J, Chandler J, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 63 (Updated February 2022): Cochrane; 2022. URL: https://www.training.cochrane.org/handbook/archive/v6.3 [accessed 2026-01-02]
29. Covidence. URL: https://www.covidence.org/ [accessed 2026-01-02]
30. Zongpa TC, Chandrasekaran B, Arumugam A. Effectiveness of a smartphone directed physical activity program on cardiometabolic disease risk in desk based office employees. a pragmatic, two arm, parallel, cluster randomised trial. Muscles Ligaments Tendons J 2020;10(4):713-723. [doi: 10.32098/mltj.04.2020.19]
31. Muntaner-Mas A, Sanchez-Azanza VA, Ortega FB, et al. The effects of a physical activity intervention based on a fatness and fitness smartphone app for University students. Health Informatics J 2021;27(1):1460458220987275. [doi: 10.1177/1460458220987275] [Medline: 33446036]
32. RoB 2: a revised Cochrane risk-of-bias tool for randomized trials. Cochraine Bias. URL: https://methods.cochrane.org/bias/resources/rob-2-revised-cochrane-risk-bias-tool-randomized-trials [accessed 2022-08-30]
33. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
34. Jamovi (version 2.3.28). jamovi. URL: https://www.jamovi.org [accessed 2026-01-07]
35. GRADE approach. Cochrane. URL: https://training.cochrane.org/grade-approach [accessed 2022-09-26]
36. Wilson D, Driller MW, Johnston B, Gill ND. A contactless app-based intervention to improve health behaviors in airline pilots: a randomized trial. Am J Prev Med 2023 May;64(5):666-676. [doi: 10.1016/j.amepre.2022.12.011] [Medline: 36641335]
37. Fanning J, Roberts S, Hillman CH, Mullen SP, Ritterband L, McAuley E. A smartphone "app"-delivered randomized factorial trial targeting physical activity in adults. J Behav Med 2017 Oct;40(5):712-729. [doi: 10.1007/s10865-017-9838-y] [Medline: 28255750]
38. Edney SM, Olds TS, Ryan JC, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):e51-e62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
39. Tulasiram B, Chandrasekaran B. Are smartphones better in guiding physical activity among sedentary young adults? A randomised controlled trial. Muscles Ligaments Tendons J 2021;11(1):77. [doi: 10.32098/mltj.01.2021.10]
40. Mamede A, Noordzij G, Jongerling J, Snijders M, Schop-Etman A, Denktas S. Combining web-based gamification and physical nudges with an app (MoveMore) to promote walking breaks and reduce sedentary behavior of office workers: field study. J Med Internet Res 2021 Apr 12;23(4):e19875. [doi: 10.2196/19875] [Medline: 33843593]
41. Balk-Møller NC, Poulsen SK, Larsen TM. Effect of a nine-month web- and app-based workplace intervention to promote healthy lifestyle and weight loss for employees in the social welfare and health care sector: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e108. [doi: 10.2196/jmir.6196] [Medline: 28396303]
42. Ek A, Alexandrou C, Delisle Nyström C, et al. The Smart City Active Mobile Phone Intervention (SCAMPI) study to promote physical activity through active transportation in healthy adults: a study protocol for a randomised controlled trial. BMC Public Health 2018 Jul 16;18(1):880. [doi: 10.1186/s12889-018-5658-4] [Medline: 30012116]
43. Glynn LG, Hayes PS, Casey M, 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]
44. Cai X, Qiu S, Luo D, et al. Effects of peer support and mobile application-based walking programme on physical activity and physical function in rural older adults: a cluster randomized controlled trial. Eur Geriatr Med 2022 Oct;13(5):1187-1195. [doi: 10.1007/s41999-022-00682-w] [Medline: 36001254]
45. King AC, Hekler EB, Grieco LA, et al. Effects of three motivationally targeted mobile device applications on initial physical activity and sedentary behavior change in midlife and older adults: a randomized trial. PLoS ONE 2016;11(6):e0156370. [doi: 10.1371/journal.pone.0156370] [Medline: 27352250]
46. Murawski B, Plotnikoff RC, Rayward AT, et al. Efficacy of an m-health physical activity and sleep health intervention for adults: a randomized waitlist-controlled trial. Am J Prev Med 2019 Oct;57(4):503-514. [doi: 10.1016/j.amepre.2019.05.009] [Medline: 31542128]
47. van Drongelen A, Boot CR, Hlobil H, Twisk JW, Smid T, van der Beek AJ. Evaluation of an mHealth intervention aiming to improve health-related behavior and sleep and reduce fatigue among airline pilots. Scand J Work Environ Health 2014 Nov;40(6):557-568. [doi: 10.5271/sjweh.3447] [Medline: 25121620]
48. Zhou M, Fukuoka Y, Mintz Y, et al. Evaluating machine learning-based automated personalized daily step goals delivered through a mobile phone app: randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 25;6(1):e28. [doi: 10.2196/mhealth.9117] [Medline: 29371177]

49. Baumann H, Heuel L, Bischoff LL, Wollesen B. Efficacy of individualized sensory-based mhealth interventions to improve distress coping in healthcare professionals: a multi-arm parallel-group randomized controlled trial. Sensors (Basel) 2023 Feb 19;23(4):2322. [doi: 10.3390/s23042322] [Medline: 36850920]
50. Arrogi A, Bogaerts A, Seghers J, et al. Evaluation of stAPP: a smartphone-based intervention to reduce prolonged sitting among Belgian adults. Health Promot Int 2019 Feb 1;34(1):16-27. [doi: 10.1093/heapro/dax046] [Medline: 28973149]
51. Gabbiadini A, Greitemeyer T. Fitness mobile apps positively affect attitudes, perceived behavioral control and physical activities. J Sports Med Phys Fitness 2019 Mar;59(3):407-414. [doi: 10.23736/S0022-4707.18.08260-9] [Medline: 29619794]
52. Murawski B, Plotnikoff RC, Lubans DR, et al. Examining mediators of intervention efficacy in a randomised controlled m-health trial to improve physical activity and sleep health in adults. Psychol Health 2020 Nov 1;35(11):1346-1367. [doi: 10.1080/08870446.2020.1756288]
53. Saran T, Pedrycz A, Mucha D, Mucha D. Follow-up monitoring of physical activity after rehabilitation by means of a mobile application: effectiveness of measurements in different age groups. Adv Clin Exp Med 2018 Aug;27(8):1037-1044. [doi: 10.17219/acem/69131] [Medline: 29962117]
54. Gremaud AL, Carr LJ, Simmering JE, et al. Gamifying accelerometer use increases physical activity levels of sedentary office workers. J Am Heart Assoc 2018 Jul 2;7(13):e007735. [doi: 10.1161/JAHA.117.007735] [Medline: 29967221]
55. Skvortsova A, Rodrigues TC, de Buisonjé D, et al. Increasing the effectiveness of a physical activity smartphone intervention with positive suggestions: randomized controlled trial. J Med Internet Res 2022 Mar 1;24(3):e32130. [doi: 10.17605/OSF.IO/CWJES]
56. Yoshimura E, Tajiri E, Michiwaki R, Matsumoto N, Hatamoto Y, Tanaka S. Long-term effects of the use of a step count-specific smartphone app on physical activity and weight loss: randomized controlled clinical trial. JMIR Mhealth Uhealth 2022 Oct 24;10(10):e35628. [doi: 10.2196/35628] [Medline: 36279159]
57. Martin SS, Feldman DI, Blumenthal RS, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 9;4(11):e002239. [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
58. Kim Y, Lee H, Chung ML. Living labs for a mobile app-based health program: effectiveness of a 24-week walking intervention for cardiovascular disease risk reduction among female Korean-Chinese migrant workers: a randomized controlled trial. Arch Public Health 2022 Aug 4;80(1):181. [doi: 10.1186/s13690-022-00941-z] [Medline: 35927769]
59. Plotnikoff RC, Jansson AK, Duncan MJ, et al. mHealth to support outdoor gym resistance training: the ecofit effectiveness RCT. Am J Prev Med 2023 Jun;64(6):853-864. [doi: 10.1016/j.amepre.2023.01.031] [Medline: 36804197]
60. Fukuoka Y, Haskell W, Lin F, Vittinghoff E. Short- and long-term effects of a mobile phone app in conjunction with brief in-person counseling on physical activity among physically inactive women: the mPED randomized clinical trial. JAMA Netw Open 2019 May 3;2(5):e194281. [doi: 10.1001/jamanetworkopen.2019.4281] [Medline: 31125101]
61. Spring B, Schneider K, McFadden HG, et al. Multiple behavior changes in diet and activity: a randomized controlled trial using mobile technology. Arch Intern Med 2012 May 28;172(10):789-796. [doi: 10.1001/archinternmed.2012.1044] [Medline: 22636824]
62. Shcherbina A, Hershman SG, Lazzeroni L, et al. The effect of digital physical activity interventions on daily step count: a randomised controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. Lancet Digit Health 2019 Nov;1(7):e344-e352. [doi: 10.1016/S2589-7500(19)30129-3] [Medline: 33323209]
63. Mönninghoff A, Fuchs K, Wu J, Albert J, Mayer S. The effect of a future-self avatar mobile health intervention (FutureMe) on physical activity and food purchases: randomized controlled trial. J Med Internet Res 2022 Jul 7;24(7):e32487. [doi: 10.2196/32487] [Medline: 35797104]
64. Gür F, Gür GC, Ayan V. The effect of the ERVE smartphone app on physical activity, quality of life, self-efficacy, and exercise motivation for inactive people: a randomized controlled trial. Eur J Integr Med 2020 Oct;39:101198. [doi: 10.1016/j.eujim.2020.101198]
65. Schroé H, Van Dyck D, De Paepe A, et al. Which behaviour change techniques are effective to promote physical activity and reduce sedentary behaviour in adults: a factorial randomized trial of an e- and m-health intervention. Int J Behav Nutr Phys Act 2020 Oct 7;17(1):127. [doi: 10.1186/s12966-020-01001-x] [Medline: 33028335]
66. Safran Naimark J, Madar Z, Shahar DR. The impact of a Web-based app (eBalance) in promoting healthy lifestyles: randomized controlled trial. J Med Internet Res 2015 Mar 2;17(3):e56. [doi: 10.2196/jmir.3682] [Medline: 25732936]
67. Damschroder LJ, Buis LR, McCant FA, et al. Effect of adding telephone-based brief coaching to an mHealth app (Stay Strong) for promoting physical activity among veterans: randomized controlled trial. J Med Internet Res 2020 Aug 4;22(8):e19216. [doi: 10.2196/19216] [Medline: 32687474]
68. MacPherson MM, Merry KJ, Locke SR, Jung ME. Effects of mobile health prompts on self-monitoring and exercise behaviors following a diabetes prevention program: secondary analysis from a randomized controlled trial. JMIR Mhealth Uhealth 2019 Sep 5;7(9):e12956. [doi: 10.2196/12956] [Medline: 31489842]
69. Olson R, Wipfli B, Hanson GC, et al. Weight loss maintenance among truck drivers in the SHIFT randomised controlled trial, USA. Occup Environ Med 2025 Jul 9;82(4):168-175. [doi: 10.1136/oemed-2024-109903] [Medline: 40355262]

70. Simmering JE, Polgreen LA, Francis SL, Strom AJ, Segre AM, Polgreen PM. Using a Fitbit-based walking game to improve physical activity among U.S. veterans. Mil Med 2025 Jan 16;190(1-2):194-201. [doi: 10.1093/milmed/usae280] [Medline: 38829720]
71. Mateo-Orcajada A, Vaquero-Cristóbal R, Mota J, Abenza-Cano L. Training and detraining effects of a physical activity program implemented through mobile applications in adolescents. ClinicalTrials.gov. 2023. URL: https://clinicaltrials.gov/study/NCT06164041 [accessed 2026-01-02]
72. Ryan DJ, Ross MH, Simmich J, et al. TRACK & ACT: a pragmatic randomised controlled trial exploring the comparative effectiveness of pedometers and activity trackers for changing physical activity and sedentary behaviour in inactive individuals. J Act Sedent Sleep Behav 2023 May 1;2(1):12. [doi: 10.1186/s44167-023-00018-4] [Medline: 40217539]
73. Collombon EHGM, Bolman CAW, de Bruijn GJ, Peels DA, Verboon P, Lechner L. The efficacy of online physical activity interventions with added mobile elements within adults aged 50 years and over: randomized controlled trial. Appl Psychol Health Well Being 2024 Nov;16(4):1921-1943. [doi: 10.1111/aphw.12568] [Medline: 38925643]
74. Lugade V, Torbitt M, O'Brien SR, Silsupadol P. Smartphone- and paper-based delivery of balance intervention for older adults are equally effective, enjoyable, and of high fidelity: a randomized controlled trial. Sensors (Basel) 2023 Aug 27;23(17):7451. [doi: 10.3390/s23177451] [Medline: 37687907]
75. Javed A, Kim DS, Hershman SG, et al. Personalized digital behaviour interventions increase short-term physical activity: a randomized control crossover trial substudy of the MyHeart Counts Cardiovascular Health Study. Eur Heart J Digit Health 2023 Oct;4(5):411-419. [doi: 10.1093/ehjdh/ztad047] [Medline: 37794870]
76. Ghazala S, Veluswamy SK, Ravindra S, Arena R, Myers J. Efficacy of mHealth-based workplace health promotion strategy in improving cardiorespiratory fitness in a healthcare setting: a randomized controlled study. J Occup Environ Med 2024 Dec 1;66(12):1083-1090. [doi: 10.1097/JOM.0000000000003229] [Medline: 39322289]
77. Zhang N, Zhou M, Li M, Ma G. Effects of smartphone-based remote interventions on dietary intake, physical activity, weight control, and related health benefits among the older population with overweight and obesity in China: randomized controlled trial. J Med Internet Res 2023 Apr 28;25:e41926. [doi: 10.2196/41926] [Medline: 37115608]
78. Sanchez-Trigo H, Maher C, Godino JG, Sañudo B. Effects of an mHealth physical activity intervention to prevent osteoporosis in premenopausal women. A randomized controlled trial. J Sci Med Sport 2023 Oct;26(10):545-552. [doi: 10.1016/j.jsams.2023.09.004] [Medline: 37739855]
79. Pettersson B, Lundin-Olsson L, Skelton DA, et al. Effectiveness of the safe step digital exercise program to prevent falls in older community-dwelling adults: randomized controlled trial. J Med Internet Res 2025 Mar 31;27:e67539. [doi: 10.2196/67539] [Medline: 40163860]
80. Chandrasekaran B, Rao CR, Pesola AJ, Arumugam A. Effectiveness of technology-assisted and self-directed interventions to sit less and move more among Indian desk-based office workers: a three-arm cluster randomised controlled trial (SMART-STEP trial). Appl Ergon 2025 Sep;127:104528. [doi: 10.1016/j.apergo.2025.104528] [Medline: 40199231]
81. Pomkai N, Potharin D, Widyastari DA, et al. Effectiveness of an mHealth application for physical activity promotion among Thai older adults: a randomized controlled trial. Inquiry 2024;61:469580241309869. [doi: 10.1177/00469580241309869] [Medline: 39718167]
82. Gómez-Cuesta N, Mateo-Orcajada A, Meroño L, Abenza-Cano L, Vaquero-Cristóbal R. A mobile app-based intervention improves anthropometry, body composition and fitness, regardless of previous active-inactive status: a randomized controlled trial. Front Public Health 2024;12:1380621. [doi: 10.3389/fpubh.2024.1380621] [Medline: 39193194]
83. Ortega FB, Ruiz JR, España-Romero V, et al. The International Fitness Scale (IFIS): usefulness of self-reported fitness in youth. Int J Epidemiol 2011 Jun;40(3):701-711. [doi: 10.1093/ije/dyr039] [Medline: 21441238]
84. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
85. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013 Nov 1;2(6):e000568. [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]
86. Car J, Gurol-Urganci I, de Jongh T, Vodopivec-Jamsek V, Atun R. Mobile phone messaging reminders for attendance at healthcare appointments. Cochrane Database Syst Rev 2012 Jul 11(7):CD007458. [doi: 10.1002/14651858.CD007458.pub2] [Medline: 22786507]
87. 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(12):CD007459. [doi: 10.1002/14651858.CD007459.pub2] [Medline: 23235644]
88. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
89. Bloomfield GS, Vedanthan R, Vasudevan L, Kithei A, Were M, Velazquez EJ. Mobile health for non-communicable diseases in Sub-Saharan Africa: a systematic review of the literature and strategic framework for research. Global Health 2014 Jun 13;10(1):49. [doi: 10.1186/1744-8603-10-49] [Medline: 24927745]

90. Worringham C, Rojek A, Stewart I. Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS ONE 2011 Feb 9;6(2):e14669. [doi: 10.1371/journal.pone.0014669] [Medline: 21347403]
91. Fukui K, Suzuki Y, Kaneda K, et al. Do “Stay-at-Home Exercise” videos induce behavioral changes in college students? A randomized controlled trial. Sustainability 2021 Nov 1;13(21):11600. [doi: 10.3390/su132111600]
92. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14(1):188. [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
93. Bacigalupo R, Cudd P, Littlewood C, Bissell P, Hawley MS, Buckley Woods H. Interventions employing mobile technology for overweight and obesity: an early systematic review of randomized controlled trials. Obes Rev 2013 Apr;14(4):279-291. [doi: 10.1111/obr.12006] [Medline: 23167478]
94. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
95. Free C, Phillips G, Watson L, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013;10(1):e1001363. [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
96. Banks AZ, Mentz RJ, Stebbins A, et al. Response to exercise training and outcomes in patients with heart failure and diabetes mellitus: insights from the HF-ACTION trial. J Card Fail 2016 Jul;22(7):485-491. [doi: 10.1016/j.cardfail.2015.12.007] [Medline: 26687984]
97. Bohannon RW. Minimal clinically important difference for grip strength: a systematic review. J Phys Ther Sci 2019 Jan;31(1):75-78. [doi: 10.1589/jpts.31.75] [Medline: 30774209]

Abbreviations

GRADE: Grading of Recommendations, Assessment, Development and Evaluation
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Literature Search Extension
RCT: randomized controlled trial
RoB: risk of bias
VO_2 max: maximal oxygen consumption

Edited by A Stone; submitted 25.Sep.2024; peer-reviewed by S Werts-Pelter, WJ He; revised version received 17.Nov.2025; accepted 17.Nov.2025; published 16.Jan.2026.

Please cite as:
Leuzzi G, Job M, Cosentino C, Testa R, Scafoglieri A, Testa M
Mobile Apps to Improve Health Parameters in Healthy Adults: Systematic Review
JMIR Mhealth Uhealth 2026;14:e66881
URL: https://mhealth.jmir.org/2026/1/e66881
doi:10.2196/66881

Review

Advancements in Wearable Sensor Technologies for Health Monitoring in Terms of Clinical Applications, Rehabilitation, and Disease Risk Assessment: Systematic Review

Bonsang Gu[1], MS; Hyeon Su Kim[1,2], BEng; HyunBin Kim[2], BEng; Jun-Il Yoo[3], MD, PhD

[1]Department of Biomedical Research Institute, Inha University Hospital, Incheon, Republic of Korea

[2]Program in Biomedical Science and Engineering, Inha University, Incheon, Republic of Korea

[3]Department of Orthopedic Surgery, Inha University Hospital, Inha University College of Medicine, Incheon, Incheon, Republic of Korea

Corresponding Author:
Jun-Il Yoo, MD, PhD
Department of Orthopedic Surgery
Inha University Hospital, Inha University College of Medicine
27, Inhang-ro, Jung-gu
Incheon, Incheon
Republic of Korea
Phone: 82 10 3242 4980
Email: furim@daum.net

Abstract

Background: Wearable sensor technologies such as inertial measurement units, smartwatches, and multisensor systems have emerged as valuable tools in clinical and real-world health monitoring. These devices enable continuous, noninvasive tracking of gait, mobility, and functional health across diverse populations. However, challenges remain in sensor placement standardization, data processing consistency, and real-world validation.

Objective: This systematic review aimed to evaluate recent literature on the clinical and research applications of wearable sensors. Specifically, it investigated how these technologies are used to assess mobility, predict disease risk, and support rehabilitation. It also identified limitations and proposed future research directions.

Methods: This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We searched the PubMed, Scopus, and Web of Science databases up to March 9, 2025. Inclusion criteria focused on studies using wearable sensors in clinical or real-world environments. A total of 30 eligible studies were identified for qualitative synthesis. Data extracted included study design, population characteristics, sensor type and placement, machine learning algorithms, and clinical outcomes.

Results: Of the included studies, 43% (13/30) were observational, 27% (8/30) were experimental, and 10% (3/30) were randomized controlled trials. Inertial measurement unit–based sensors were used in 67% (20/30) of the studies, with wrist-worn devices being the most common (13/20, 65%). Machine learning techniques were frequently applied, with random forest (6/30, 20%) and deep learning (5/30, 17%) models predominating. Clinical applications spanned Parkinson disease, stroke, multiple sclerosis, and frailty, with several studies (4/30, 13%) reporting high predictive accuracy for fall risk and mobility decline (area under the receiver operating characteristic curve up to 0.97).

Conclusions: Wearable sensors show strong potential for mobility monitoring, disease risk assessment, and rehabilitation tracking in clinical and real-world settings. However, challenges remain in standardizing sensor protocols and data analysis. Future research should focus on large-scale, longitudinal studies; harmonized machine learning pipelines; and integration with cloud-based health systems to improve scalability and clinical translation.

(JMIR Mhealth Uhealth 2026;14:e76084) doi:10.2196/76084

KEYWORDS

wearable sensors; health monitoring; systematic review; gait analysis; rehabilitation; mobile health; mHealth

Introduction

Throughout this paper, we use standardized terminology to ensure clarity and consistency. The term "wearable sensor" refers to any body-worn device capable of measuring physiological or biomechanical parameters, including inertial measurement unit (IMU)–based sensors, smartwatches, smart insoles, and multisensor systems. "IMU" refers specifically to devices incorporating accelerometers, gyroscopes, and magnetometers for motion tracking. When referring to specific device types (eg, smartwatches and smart insoles), we use the precise terminology to distinguish their unique features and applications.

Wearable sensors have gained significant attention in clinical research and health care for their ability to provide continuous, real-world assessments of mobility and physiological health. These devices, including IMU-based sensors, smartwatches, and multisensor systems, have transformed traditional gait and activity monitoring by enabling remote, noninvasive tracking of movement patterns and health status [1]. The integration of advanced analytics, particularly machine learning (ML), has further enhanced their diagnostic and predictive capabilities, positioning wearable sensors as key tools in digital health and precision medicine [2].

Gait analysis and mobility tracking have been central to wearable sensor applications, particularly in neurological, musculoskeletal, and age-related conditions. In Parkinson disease (PD), wearable sensors have been used to detect subtle changes in gait speed, stance and swing phase durations, and postural instability, aiding in early disease detection and progression monitoring. In stroke rehabilitation, these sensors enable remote motor recovery assessment and provide continuous mobility data outside traditional clinical settings [3]. Wearable sensors also demonstrate high efficacy in frailty assessment and fall risk prediction, offering objective, real-time alternatives to conventional tools such as the Performance-Oriented Mobility Assessment (POMA) and Timed Up and Go (TUG) tests.

Despite their growing clinical adoption, several challenges hinder the widespread implementation of wearable sensor technology. Variability in sensor placement, study methodologies, and data processing techniques limits cross-study comparability and reproducibility [3]. Additionally, while controlled laboratory studies have validated their accuracy, real-world validation remains insufficient, necessitating further large-scale, longitudinal studies to assess their usability and reliability across diverse populations [4]. Furthermore, standardization of ML frameworks and data interpretation methodologies is essential to ensure consistent clinical application [5].

This systematic review aimed to provide a comprehensive evaluation of wearable sensor research, analyzing their clinical applications, technological advancements, and methodological challenges. By synthesizing evidence from recent studies, we highlight key trends in wearable sensor use, discuss their implications for health care, and propose future directions to enhance their impact in mobility monitoring and rehabilitation.

Methods

Study Design

The protocol of this review was not registered in PROSPERO due to its exploratory nature and inclusion of emerging sensor studies. However, the review process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidance to ensure methodological transparency. The study selection process followed PRISMA 2020 guidelines, including identification, screening, eligibility assessment, and final inclusion. This review focused on studies published in the last 10 years that investigated the applications and effectiveness of wearable sensors, including smartwatches, in remote health monitoring, rehabilitation, and disease assessment. Full-text articles were included to ensure a comprehensive analysis. We aimed to synthesize evidence on clinical and research applications of wearable sensors, particularly for gait analysis, fall risk assessment, and disease monitoring. Given study heterogeneity, we categorized and synthesized the findings narratively, emphasizing disease-specific insights and sensor use trends.

Search Strategy

A comprehensive database search was conducted across PubMed, Scopus, and Web of Science.

The search strategy combined terms related to wearable technologies, inertial sensors, digital biomarkers, and rehabilitation. Representative Boolean operators were used as follows:

> *("smartwatch" OR "smart watch" OR "wearable sensor" OR "wearable sensor") AND ("accelerometer" OR "acceleration sensor" OR "inertial sensor" OR "IMU") AND ("remote monitoring" OR "digital biomarkers" OR "telemedicine" OR "wearable health tracking") AND (aging OR older adults OR elderly OR Parkinson OR stroke OR "gait disorders" OR "neurological disorders" OR "movement disorders" OR "fall risk" OR rehabilitation OR "functional mobility" OR sarcopenia OR osteoarthritis OR dementia) NOT review.*

The initial search yielded 4226 records. Of these 4226 records, after removing duplicates and studies unrelated to clinical applications (n=3664, 86.7%), 562 (13.3%) remained for screening. Of these 562 studies, those focusing solely on technical performance comparisons (n=501, 89.1%) were excluded, leaving 61 (10.9%) for eligibility assessment. Of these 61 studies, an additional 31 (51%) were excluded due to limited relevance to disease-related applications, resulting in 30 (49%) studies included in the final review. The search was finalized on March 9, 2025. In total, 30 studies met the inclusion criteria and were included in the final synthesis. Additional references cited throughout the manuscript (n=43) were used for background, context, and methodological justification.

Study Selection Process

A single researcher conducted study selection and data extraction following a predefined protocol to minimize bias. The eligibility criteria were clearly defined and consistently applied. Studies published from March 9, 2015, to March 9, 2025, were considered, reflecting a 10-year search window. Eligible study designs included randomized controlled trials (RCTs), observational studies, and experimental validation studies conducted in either clinical or real-world settings. Both research-grade and commercial wearable sensors were included provided that they reported measurable health or functional outcomes. Only English-language, peer-reviewed articles were included, and conference abstracts, reviews, and purely technical feasibility reports without human participants were excluded. Quality appraisal using the Newcastle-Ottawa Scale (NOS), Joanna Briggs Institute (JBI) appraisal tools, or version 2 of the Cochrane risk-of-bias tool for randomized trials (RoB 2) was conducted to describe study rigor but did not influence inclusion decisions.

Any uncertainties during the selection process were resolved by re-evaluating studies against the predefined inclusion criteria. Data extraction was conducted manually using a standardized form. No independent reviewers cross-checked the extracted data, which is acknowledged as a limitation. Missing or unclear data were clarified when possible, by contacting the corresponding authors. No automation tools were used for data collection.

Participant Selection in the Included Studies

The included studies targeted diverse populations, including healthy adults; older individuals; and patients diagnosed with neurological disorders (eg, stroke and PD), musculoskeletal disorders (eg, sarcopenia and osteoarthritis), or metabolic conditions (eg, diabetes). Participants could walk independently and provided informed consent. We excluded studies lacking clear participant definitions or standard gait analysis metrics to maintain consistency. Pediatric studies were excluded except for those including toddler cohorts (aged <3 years) and specifically designed for developmental gait analysis. To ensure data quality, studies were required to report a minimum wear time of 30 minutes of valid sensor data per session.

Wearable Sensor Technology

The wearable sensors reviewed featured advanced components such as high-precision accelerometers, gyroscopes, and pressure sensors. These sensors accurately captured key gait and mobility parameters: step length, stance and swing phase durations, plantar pressure distribution, and center of pressure. Smartwatches were primarily used for activity tracking and remote monitoring, whereas wearable sensors and foot-mounted sensors specialized in gait and postural assessments. The wearable sensors integrated seamlessly into daily life, ensuring high usability and real-world applicability.

Data Acquisition and Analysis

Data were collected using a variety of wearable sensor systems, including IMUs, smart insoles, smartwatches, and pressure-sensing devices. Sensor placement varied by study objective and included the wrist, waist, ankle, thigh, lumbar spine, and foot. The IMU sensors incorporated accelerometers, gyroscopes, and magnetometers with sampling rates ranging from 10 to 1149 Hz depending on the device and measurement context. Pressure-sensitive insoles provided additional biomechanical insights through plantar pressure distribution and force-time characteristics.

Wireless data transmission via Bluetooth or cloud platforms enabled real-time monitoring and digital biomarker extraction. Embedded preprocessing algorithms were applied to reduce noise, improve signal quality, and enhance feature extraction accuracy. Studies used various ML techniques, including random forest (6/30, 20%), deep learning (5/30, 17%), elastic net regression (4/30, 13%), and principal component analysis (PCA; 2/30, 7%), for pattern recognition, mobility classification, and disease risk prediction. All reported quantitative values (eg, area under the receiver operating characteristic curve [AUROC], accuracy, and improvement rate) were extracted from individual studies and are presented descriptively, not as pooled estimates.

The methodological quality of the included studies was systematically evaluated using appropriate assessment tools based on the study design. The NOS was applied to prospective cohort studies, whereas the JBI critical appraisal checklist was used for observational, cross-sectional, and experimental studies. For RCTs, the RoB 2 tool was used to ensure a robust evaluation of study quality. The results of the quality assessment guided the interpretation of the reliability and clinical applicability of the findings. Studies were categorized as low (JBI or NOS score of ≥8), moderate (JBI or NOS score of 6-7), or high (JBI or NOS score of ≤5) risk of bias according to established thresholds.

Ethical Considerations

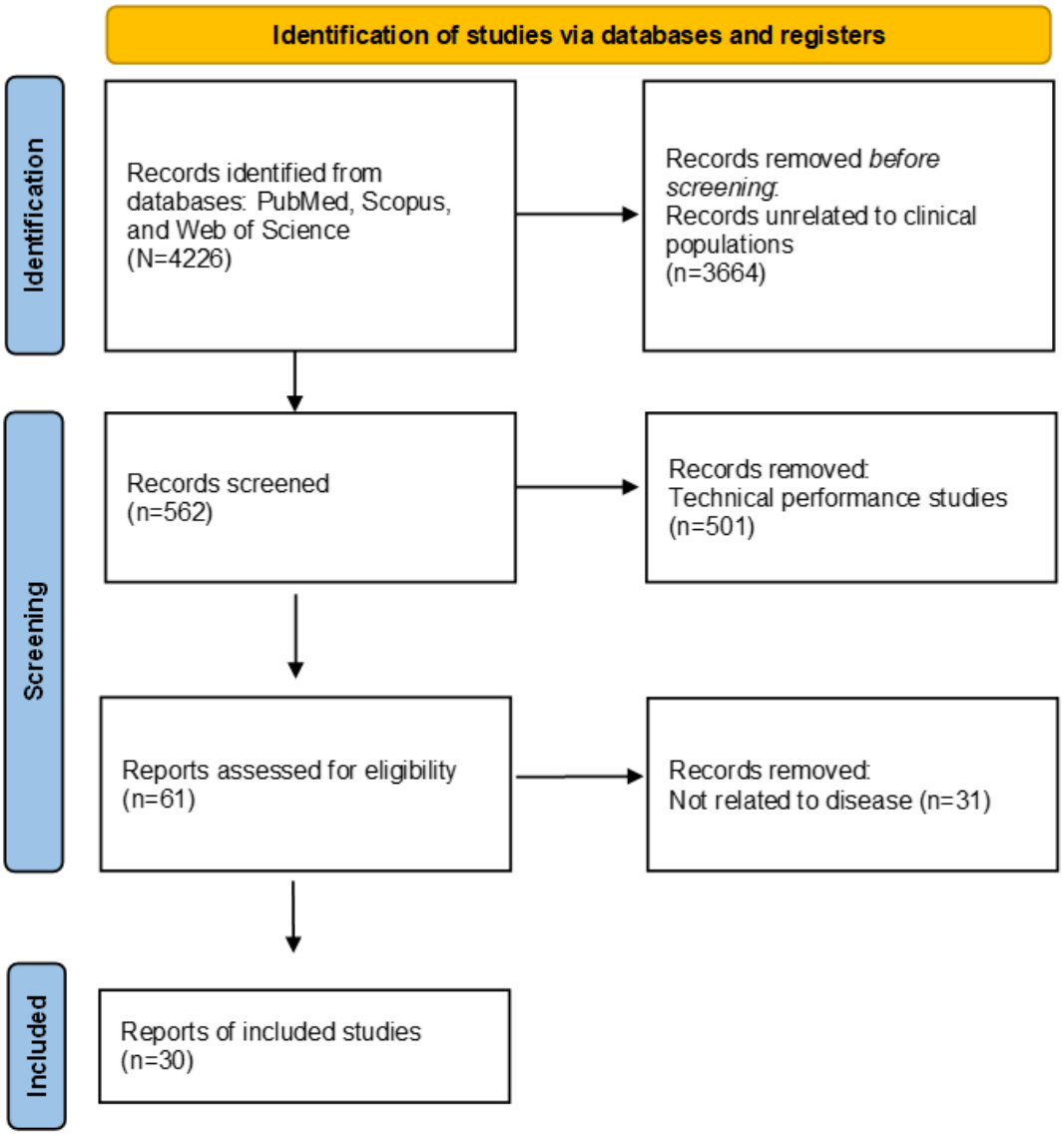
All studies included adhered to ethical standards for human research, following the Declaration of Helsinki. As this study is a systematic review of published literature and did not involve human participants, interventions, or identifiable private data, ethics approval was not required. Data privacy and participant well-being were prioritized across the studies.

Results

Study Design and Population Characteristics

A total of 4226 records were identified through database searching. After removing duplicates and screening titles and abstracts, 562 articles remained for full-text assessment. Of these, 30 studies met the inclusion criteria and were included in the final review. The full screening process is summarized in Figure 1 (PRISMA 2020 flow diagram). Among the analyzed studies, the most frequently used research design was prospective observational studies, accounting for 43% (13/30) of the total. Experimental studies comprised 27% (8/30), whereas RCTs were limited to 10% (3/30). Cross-sectional studies and cohort studies accounted for 13% (4/30) and 7% (2/30), respectively. Observational studies were most common due to their feasibility, whereas experimental studies validated sensor-based assessments. RCTs were scarce, indicating limited rigorous intervention-based evaluations. Cohort studies were included in long-term monitoring applications.

Figure 1. PRISMA flow diagram illustrating the screening process of papers for study selection.



The studies focused on various populations, with healthy adults being the most common participant group, as summarized in Table 1. Studies involving neurodegenerative diseases such as PD, stroke, Huntington disease, and multiple sclerosis (MS) were also prevalent. Healthy adults were often included for sensor validation and reliability assessment, whereas studies on neurodegenerative diseases primarily aimed at mobility and functional monitoring. Research on frailty in older adults (4/30, 13%) focused on mobility assessment, balance, and fall risk analysis. The reviewed studies covered a wide range of age groups and health conditions. Studies involving healthy adults typically included participants from early adulthood to middle age, with some extending to adolescent and pediatric populations. Research on neurodegenerative diseases such as PD and Huntington disease focused primarily on older adults, whereas cardiovascular disease– and frailty-related studies involved older participants as well. Certain studies (12/30, 40%) targeted specific conditions such as osteoarthritis, rheumatoid arthritis, MS, cystic fibrosis, stroke, and spinal cord injury, highlighting the diverse application of wearable sensors, particularly smartwatches, in different clinical populations. The distribution of male and female participants varied across studies, with some (1/30, 3%) focusing specifically on age-based differences in sensor performance and health monitoring.

Table 1. Summary of study populations in the reviewed studies.

Study	Participant age (y)	Number of male/female participants	Health condition
Bolam et al [6], 2021	Mean 66.8 (SD 7.0)	6/8	Knee osteoarthritis
Angelucci and Aliverti [7], 2023	Mean 26.8 (SD NR[a]; range 23-54)	9/11	Healthy adults
Greene et al [8], 2021	PD[b] 1 (clinical cohort): mean 67.3 (SD 7.1); PD 2 (exercise cohort): mean 64.9 (SD 7.3)	27/17	PD
Gordon et al [9], 2019	Mean 51 (SD 12)	9/8	HD[c]
Presley et al [10], 2023	Range 13-24	6/12	Healthy adults and adolescents
De Cannière et al [11], 2020	Mean 63 (SD 1)	65/24	CVD[d] and heart failure
Nunes et al [12], 2024	HD: mean 51.9 (SD 11); pre-HD stage: mean 36.5 (SD 13.1); control: mean 58.9 (SD 12.2)	8/8 (HD group)	HD, pre-HD stage, and controls
Mahadevan et al [13], 2020	Control: mean 43.9 (SD 10); PD: mean 68.1 (SD 8.1)	Control: 27/33; PD: 23/12	PD
Seo et al [14], 2024	Mean 61 (SD 12)	12/7	Stroke (upper-limb hemiparesis)
Odhiambo et al [15], 2023	Mean 27.25 (SD NR; range 20-56)	16/12	Healthy adults
Hwang and Effenberg [16], 2021	Mean 29.8 (SD 6.8)	6/6	Healthy adults
Wu et al [17], 2021	Mean 48 (range 26-66)	2/2	Healthy adults
John and Soangra [18], 2022	Control: mean 74 (SD 8.7); Stroke: mean 69 (SD 8.4)	Control: 6/8; Stroke: 8/6	Stroke vs healthy older adults
Meyer et al [19], 2022	Mean 51 (SD 7)	5/16	Multiple sclerosis
Toumieux et al [20], 2015	Not specified	Not specified	Stroke
Elstub et al [21], 2022	Not specified	6/3	Healthy runners
Perraudin et al [22], 2018	RA[e]: mean 50.7 (SD 11.4); PA[f]: mean 47.5 (SD 15.5); OA[g]: mean 60.7 (SD 4.5); healthy: mean 48 (SD 13.6)	RA: 18 female; PA: 2 female; OA: 10 female; healthy: 15 female	RA, PA, and OA
Giggins et al [23], 2025	Mean 77.5 (SD 8.4)	12/39	Frailty levels
Savi et al [24], 2020	Mean 37.5 (SD 11.5)	6/18	Cystic fibrosis
Haghi et al [25], 2023	Mean 26.7 (SD NR)	22/8	Healthy adults
DasMahapatra et al [26], 2018	Mean 52 (SD 10.6)	28/86	Multiple sclerosis
Sun et al [27], 2019	OA: mean 78.2 (SD 6.1); YA[h]: mean 24.4 (SD 3.9)	OA: 4/6; YA: 6/6	Healthy adults
Ramezani et al [28], 2019	Community: mean 82.16 (SD 9.55); hospital: mean 84.22 (SD 13.87)	Community: 41/104; hospital: 5/4	Older adults with multiple chronic conditions
Liew et al [29], 2024	Mean 20.8 (SD 1.6)	9/9	Healthy adults
Kwon et al [30], 2019	13-35 months (1-year-olds: 11; 2-year-olds: 11)	10/12	Healthy toddlers
Martin et al [31], 2015	Mean 58 (SD 8; range 18-69)	26/22	CVD prevention and obesity
Hup et al [32], 2024	Various	Not specified	OHCA[i] and SCA[j]
Browne et al [33], 2020	Mean 13.3 (SD 2.3)	9/11	Pediatric obesity
Burns et al [34], 2020	≥18	Not specified	Rotator cuff and shoulder pain
Bailey et al [35], 2024	Mean 51 (SD 9; range 28-63)	11/4	Spinal cord injury and wheelchair users

[a]NR: not reported.

[b]PD: Parkinson disease.

[c]HD: Huntington disease.
[d]CVD: cardiovascular disease.
[e]RA: rheumatoid arthritis.
[f]PA: psoriatic arthritis.
[g]OA: old adult.
[h]YA: young adult.
[i]OHCA: out-of-hospital cardiac arrest.
[j]SCA: sudden cardiac arrest.

Sensor Use and Data Analysis in Wearable Research

Studies relied heavily on IMU sensors (20/30, 67%) and smartwatches (8/30, 27%). Shoe-mounted sensors and multisensor systems incorporating electrocardiograms were less frequently used. Wrist-worn sensors (13/30, 43%) were the most common due to ease of wear and practical data collection. Smartwatches, as a subset of wrist-worn devices, were frequently used for continuous activity tracking and health monitoring. Additionally, ankle and thigh placements (7/30, 23%) were primarily used for gait analysis, whereas foot and insole sensors (2/30, 7%) were implemented for more specialized balance and gait assessments.

The predominant activity type studied was gait analysis, which appeared in 60% (18/30) of the studies, followed by activities of daily living (12/30, 40%), balance assessments (8/30, 27%), energy expenditure evaluations (6/30, 20%), and rehabilitation exercises (6/30, 20%). Gait analysis was especially relevant in research focused on neurodegenerative diseases and mobility impairments, where smartwatches were often used for free-living gait monitoring. Studies on activities of daily living leveraged smartwatches for continuous data collection in real-world settings. Balance assessments were primarily conducted for frailty and fall risk evaluations, with some smartwatch-based apps integrating accelerometry for postural control analysis.

Data processing in wearable sensor research used a range of ML techniques. Random forest was the most commonly applied method (6/30, 20%), followed by deep learning models (5/30, 17%), elastic net regression and support vector machines (SVMs; 4/30, 13%), and PCA (2/30, 7%). Random forest was frequently used in gait analysis and activity recognition, whereas deep learning models were applied for long-term movement pattern analysis, particularly in smartwatch-based apps. Elastic net and SVM were commonly used for classification tasks, and PCA was used for dimensionality reduction, optimizing the performance of wearable sensor data processing.

Clinical Applications of Wearable Sensors

Overview

The diverse clinical applications of wearable sensors, categorized into 5 main areas—healthy individuals, age-related conditions, neurological conditions, musculoskeletal disorders, and metabolic conditions—are summarized in Table 2.

Table 2. Summary of clinical results for wearable sensors across populations.

Category and condition	Key findings	Clinical significance and utility
Rehabilitation assessment and functional recovery		
Knee arthroplasty (TKA[a] and UKA[b])	Bone stimulus: +52%; impact load: +371%; OKS[c]: +52%; EQ-5D: +32%	IMU[d]-based wearables support accurate postoperative monitoring and personalized recovery for knee arthroplasty.
CR[e]	6MWD[f] prediction error: 42.8 m; R^2=0.661	CR progress can be reliably tracked remotely, improving long-term care.
Stroke—upper-limb rehabilitation	Movement quality classification accuracy: 92%; F_1-score: 0.95	IMUs enable remote monitoring of home-based rehabilitation, enhancing adherence and personalization.
Rotator cuff injury rehabilitation	Exercise classification accuracy: 99.99%	Smartwatches improve rehabilitation compliance and exercise tracking at home.
Lumbar mobility assessment	Significant ROM[g] differences between wrist and lumbar sensors (up to 8, *P*=.003)	Wrist-worn sensors allow for remote lumbar mobility assessment for rehabilitation use.
Disease state prediction and risk assessment		
PD[h]—tremor and fall risk	Tremor detection accuracy: 83%; sensitivity: 86%; specificity: 86%; fall risk prediction RMSE[i]: 0.42	Wearables objectively monitor PD symptoms and fall risk for early intervention.
HD[j]	Sensitivity: 85%; specificity: 72%; accuracy: 81%; AUROC[k]: 0.82	HD motor decline can be tracked remotely for personalized care.
Frailty assessment in older adults	QTUG[l] accuracy: 75.8%; ScanWatch-enhanced model: 79.3%	Wearables detect frailty early, enabling preventive intervention in older adults.
OHCA[m] detection	Optimized balance between sensitivity and specificity	OHCA can be detected in real time via wearables for rapid emergency response.
Activity and behavior tracking		
Arthritis—pain and function monitoring	5-STS[n] performance significantly correlated with morning pain scores (*P*<.05)	IMUs enable noninvasive, remote tracking of arthritis pain and function.
CF[o]—activity monitoring	Fitbit and iOS smartphone showed strong agreement with SWA[p]	Consumer wearables offer scalable, affordable physical activity monitoring.
Consumer vs research-grade wearables (activity monitoring)	Z-Track sedentary behavior detection: AUC[q]=0.95; MVPA[r] detection: AUC=0.93	Fitbit-level devices reliably track sedentary behavior and activity levels.
Medication adherence monitoring	Medication intake detection accuracy: 93.6%; sensitivity: 92%	Wearables automate medication tracking, improving adherence in chronic care.
Gait analysis and balance assessment		
Gait symmetry analysis (head-worn sensor)	Gait event detection accuracy: 99.35%	Head-worn sensors support gait symmetry analysis for neurological rehabilitation.
Balance assessment (smartwatch based)	Strong correlation between smartwatch and research-grade sensors (*r*=0.861-0.970)	Smartwatches enable balance monitoring at home to prevent falls.
MS[s]—balance and mobility	AUROC=0.97	Wearables track long-term mobility and balance in MS, supporting personalized rehabilitation.

[a]TKA: total knee arthroplasty.

[b]UKA: unicompartmental knee arthroplasty.

[c]OKS: Oxford knee score.

[d]IMU: inertial measurement unit.

[e]CR: cardiac rehabilitation.

[f]6MWD: 6-minute walk distance.

[g]ROM: range of motion.

[h]PD: Parkinson disease.

[i]RMSE: root mean square error.

[j]HD: Huntington disease.

[k]AUROC: area under the receiver operating characteristic curve.

[l]QTUG: quantitative timed up and go.

[m]OHCA: out-of-hospital cardiac arrest.

[n]5-STS: 5-time sit-to-stand assessment.

[o]CF: cystic fibrosis.

[p]SWA: sensewear armband.

[q]AUC: area under the curve.

[r]MVPA: moderate to vigorous physical activity.

[s]MS: multiple sclerosis.

Rehabilitation Assessment and Functional Recovery

Wearable sensors were used to analyze gait metrics in both young and older adults. Studies on young adults focused on plantar pressure distribution, step length, swing time, and ground reaction force, achieving high accuracy in real-time gait analysis, such as 95% using the FreeWalker system with a 1000-Hz sampling frequency. Advanced ML techniques further enhanced center of pressure prediction accuracy by over 30%. Among older adults, wearable sensors were effective in assessing fall risk and mobility. Improvements were observed in swing time (+6.45%) and slip and trip classification accuracy, which exceeded 98% (P<.05).

Disease State Prediction and Risk Assessment

Studies addressed frailty and fall history using wearable sensors to measure load distribution, gait phases, and stance and swing time. Load distribution assessments demonstrated high reliability, with intraclass correlation coefficient values reaching 0.91 and strong correlations for the left (r=0.7171) and right (r=0.7502) foot. Fall risk indexes provided significant predictive accuracy, with AUROC values of 0.919 (P<.05), making them comparable to traditional tools such as the POMA and TUG tests. These findings emphasize the potential of wearable sensors for early identification of frailty and fall risk in older adults.

Activity and Behavior Tracking

Wearable sensors were used to evaluate gait characteristics in individuals with stroke, MS, and PD. Among stroke survivors, significant reductions in gait speed and step length were observed compared to controls, with strong correlations between Fugl-Meyer Assessment lower-limb scores and stance time differences (R^2=0.71). In MS, high agreement was reported between the FeetMe and GAITRite systems (intraclass correlation coefficient>0.8), validating the utility of wearable sensors for mobility monitoring. For individuals with PD, significant differences were detected in gait speed, stride length, and swing and stance time compared to healthy controls (P<.05), demonstrating the role of wearable sensors in tracking disease progression.

Gait Analysis and Balance Assessment

Wearable sensors were effective in managing diabetes and other metabolic disorders. For diabetes, total contact casts reduced forefoot contact area by 5% and peak pressure by 8% (P<.05), effectively offloading pressure and reducing the risk of complications. These devices provide actionable data that support better management of metabolic health and reduce disease-related complications.

Quality Assessment Results

The quality assessment results are summarized in Table 3, with the studies rated using the JBI critical appraisal tools scoring between 5 and 8 out of 10, NOS-rated cohort studies scoring between 6 and 7 out of 9, and RoB 2–rated RCTs scoring 8 out of 10. Of the 30 included studies, 6 (20%) were rated as low risk, and 24 (80%) were rated as having a moderate risk of bias. Studies investigating wearable sensor–based mobility assessments, gait analysis, and rehabilitation applications showed high feasibility and reliability, particularly those incorporating real-time monitoring and signal processing techniques. However, several studies (26/30, 87%) exhibited limitations such as small sample sizes, lack of validation in real-world settings, and limited applicability to diverse patient populations. Additionally, some studies (12/30, 40%) faced technical challenges, including sensor displacement errors, signal-to-noise ratio issues, and data synchronization difficulties.

Table 3. Quality assessment summary of the reviewed studies.

Study	Study design	Quality score	Risk-of-bias category
Bolam et al [6], 2021	Prospective cohort study	7/9 (NOS[a])	Moderate
Angelucci and Aliverti [7], 2023	Experimental study	6/10 (JBI[b] tool)	Moderate
Greene et al [8], 2021	Experimental study	7/10 (JBI tool)	Moderate
Gordon et al [9], 2019	Observational study	6/10 (JBI tool)	Moderate
Presley et al [10], 2023	Experimental study	8/10 (JBI tool)	Low
De Cannière et al [11], 2020	Prospective cohort study	7/9 (NOS)	Moderate
Nunes et al [12], 2024	Observational study	6/10 (JBI tool)	Moderate
Mahadevan et al [13], 2020	Observational study	8/10 (JBI tool)	Low
Seo et al [14], 2024	Observational study	7/10 (JBI tool)	Moderate
Odhiambo et al [15], 2023	Experimental study	8/10 (JBI tool)	Low
Hwang and Effenberg [16], 2021	Observational study	7/10 (JBI tool)	Moderate
Wu et al [17], 2021	Observational study	7/10 (JBI tool)	Moderate
John and Soangra [18], 2022	Observational study	6/10 (JBI tool)	Moderate
Meyer et al [19], 2022	Observational study	7/10 (JBI tool)	Moderate
Toumieux et al [20], 2015	Experimental study	5/10 (JBI tool)	Moderate
Elstub et al [21], 2022	Experimental study	7/10 (JBI tool)	Moderate
Perraudin et al [22], 2018	Observational study	7/10 (JBI tool)	Low
Giggins et al [23], 2025	Cross-sectional study	6/10 (JBI tool)	Moderate
Savi et al [24], 2020	Cross-sectional study	7/10 (JBI tool)	Moderate
Haghi et al [25], 2023	Experimental study	8/10 (JBI tool)	Moderate
DasMahapatra et al [26], 2018	Observational study	6/10 (JBI tool)	Low
Sun et al [27], 2019	Experimental study	7/10 (JBI tool)	Moderate
Ramezani et al [28], 2019	Pilot study	6/10 (JBI tool)	Moderate
Liew et al [29], 2024	Cross-sectional study	6/10 (JBI tool)	Moderate
Kwon et al [30], 2019	Observational study	7/10 (JBI tool)	Moderate
Martin et al [31], 2015	RCT[c]	8/10 (RoB 2[d])	Low
Hup et al [32], 2024	Clinical study	7/10 (JBI tool)	Moderate
Browne et al [33], 2020	RCT	8/10 (RoB 2)	Moderate
Burns et al [34], 2020	Prospective cohort study	7/9 (NOS)	Moderate
Bailey et al [35], 2024	Cross-sectional study	7/10 (JBI tool)	Moderate

[a]NOS: Newcastle-Ottawa Scale.

[b]JBI: Joanna Briggs Institute.

[c]RCT: randomized controlled trial.

[d]RoB 2: version 2 of the Cochrane risk-of-bias tool for randomized trials.

Discussion

Expanding the Role of Wearable Sensors in Health Monitoring

Wearable sensor technology, including IMU-based smartwatches, smart insoles, and multisensor systems, has significantly transformed health monitoring, rehabilitation tracking, and disease risk assessment [36]. These devices enable continuous, real-world tracking of mobility and functional health, addressing key limitations of traditional clinical assessments [37]. The reviewed studies highlight these devices' diverse applications in neurological, musculoskeletal, cardiovascular, and metabolic conditions, supporting early disease detection, remote therapy adherence, and precision rehabilitation [38].

Observational studies accounted for 43% (13/30) of the reviewed studies, reflecting the feasibility of longitudinal monitoring, whereas experimental studies made up 27% (8/30), playing a crucial role in validating sensor-based assessments. However,

the limited number of RCTs, representing only 10% (3/10) of the studies, underscores the need for rigorous intervention-based research to establish causal relationships between wearable sensor use and patient outcomes. The study populations were diverse, with healthy individuals comprising 33.3% of the participants, often included for sensor validation and reliability testing. Clinical populations, including individuals with PD, stroke, frailty, and cardiovascular conditions, were the primary focus of applied research, demonstrating the potential for wearable sensors to support patient management in real-world health care settings.

Evolution of Wearable Sensor Applications

In our review, 30% (9/30) of the included studies conducted validation in healthy adults, indicating that early wearable-sensor research primarily focused on device feasibility and performance testing before expanding into clinical populations. These devices have revolutionized mobility monitoring, particularly in neurodegenerative conditions such as PD, MS, and stroke, where continuous tracking of gait parameters enables early detection of motor impairments and disease progression. They also play a significant role in frailty assessment (4/30, 13%) and fall risk prediction. Smart insoles demonstrate high predictive accuracy (AUROC=0.919; P<.05) as noninvasive, real-world mobility assessment tools.

Technological Integration and Advances in Data Processing

The studies primarily used IMU-based systems (20/30, 67%) and smartwatches (8/30, 27%). Wrist-worn sensors were the most common, representing 43% (13/30) of the devices used, as they offer practicality, ease of wear, and convenience for everyday use. Ankle- and thigh-mounted sensors accounted for 23% (7/30) of applications and were primarily used for gait and posture assessments, whereas multisensor systems integrating electrocardiograms and pressure sensors provided additional biomechanical and cardiovascular insights, although they were less frequently studied.

Advances in ML have significantly enhanced data interpretation and predictive capabilities in wearable sensor applications [2]. Random forest models, applied in 20% (6/30) of the studies, were widely used for gait classification and activity recognition, whereas deep learning techniques were applied in 17% (5/30) of the studies and demonstrated high accuracy in long-term movement analysis. Elastic net regression and SVMs were used in 13% (4/30) of cases for classification tasks, whereas PCA was used in 7% (2/30) of the studies to reduce dimensionality and optimize data processing. However, variability in feature extraction methods remains a challenge. Standardized approaches are needed to improve reproducibility and clinical translation.

Clinical Applications of Wearable Sensors

Wearable sensors demonstrated strong feasibility across multiple health care applications, including rehabilitation monitoring, disease risk assessment, activity tracking, and gait analysis. For rehabilitation assessment, wearable sensors improved postsurgical monitoring in patients who underwent knee arthroplasty, showing 52% better bone stimulus and 371% better impact load tracking. Wearable sensors for cardiac rehabilitation demonstrated reliable 6-minute walk distance prediction, with an error of 42.8 m and an R^2 value of 0.661, facilitating remote patient monitoring. In stroke rehabilitation, IMU-based movement quality assessments achieved 92% accuracy (F_1-score=0.95), supporting their use for personalized therapy and remote monitoring.

Recent studies have also extended the application of IMU-based wearable sensors to shoulder rehabilitation. Tranquilli et al [39] demonstrated that a single IMU could simultaneously capture joint mobility and muscle strength dynamics during postinjury recovery. Ajčević et al [40] applied IMU sensors to quantify shoulder kinematics and evaluate therapeutic response in adhesive capsulitis, whereas Parel et al [41] introduced a kinematic biofeedback program integrating inertial sensors for patients after rotator cuff repair. These studies highlight the versatility of IMU technology for upper-limb functional assessment and real-time feedback during rehabilitation.

Wearable sensors also played a key role in disease prediction and risk assessment. In PD monitoring, wearable technology achieved an accuracy of 83% in tremor detection and both a sensitivity and specificity of 86% in fall risk prediction, supporting the feasibility of early intervention strategies. Fall risk assessments using wearable sensors reached an AUROC value of 0.919, demonstrating their ability to provide noninvasive, real-world alternatives to clinical assessments such as the TUG and POMA tests.

Activity and behavior tracking applications showed high accuracy, particularly in arthritis-related pain and function monitoring, where significant correlations were observed between morning pain scores and 5-time sit-to-stand performance, with P values of less than .05. Consumer wearables such as Fitbit and iOS-integrated smartwatches achieved a strong agreement with research-grade sensors, with an AUROC of 0.93, demonstrating their feasibility for large-scale, real-world activity tracking. In medication adherence monitoring, smartwatch-based tracking achieved an accuracy of 93.6%, highlighting its potential for improving adherence in chronic disease management.

Gait and balance assessments using wearable sensors provided highly accurate insights into functional mobility. Head-worn IMU–based gait symmetry analysis reached an accuracy of 99.35%, indicating its effectiveness in neuromuscular rehabilitation and postural correction. Wearable sensor–based assessments of balance and mobility for patients with MS achieved an AUROC of 0.97, reinforcing their potential to support personalized rehabilitation planning and disease progression monitoring.

The included studies encompassed diverse populations, including healthy adults, neurological patients (eg, PD), individuals with musculoskeletal disorders, and pediatric or rehabilitation cohorts. This diversity introduces biomechanical and physiological variability in gait patterns, movement strategies, and sensor placement feasibility. Differences in muscle coordination, assistive device use, and experimental environments further contribute to heterogeneity. Given these

variations, direct quantitative comparisons between studies were avoided. Instead, a narrative synthesis was used to identify overarching technological and methodological trends across populations. This approach emphasizes generalizable insights—such as the importance of standardized placement, calibration, and cross-population validation—while acknowledging disease-specific distinctions in biomechanics and sensor performance.

Many of the included studies (19/30, 63%) used ML algorithms such as random forest, deep learning, elastic net regression, and PCA for signal interpretation and disease classification. However, reporting transparency and methodological rigor varied substantially. Several studies (26/30, 87%) were limited by small sample sizes and internal validation only, increasing the risk of model overfitting. In addition, few studies (4/30, 13%) provided sufficient details regarding cross-validation protocols, feature selection strategies, or hyperparameter optimization.

Adherence to standardized ML reporting frameworks—such as the Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis–Artificial Intelligence and Prediction of Model Risk of Bias Assessment Tool–Artificial Intelligence—was rarely observed, which may affect reproducibility and generalizability.

Future research should emphasize external validation, open-source code sharing, and adherence to established ML reporting standards to ensure reliability and transparency in sensor-based clinical modeling.

Challenges in Wearable Sensor Research

Despite the promising applications of wearable sensors, several challenges remain that must be addressed to ensure widespread clinical adoption and real-world impact.

First, small sample sizes (26/30, 87% of the studies) and limited real-world validation (12/30, 40% of the studies) reduced finding generalizability. Short study durations (8/30, 27%) also hindered long-term effectiveness assessment. Beyond these methodological limitations, the scope of this review was restricted to English-language, peer-reviewed publications, excluding gray literature such as conference abstracts and theses. This language restriction and publication bias may have favored studies reporting positive or statistically significant outcomes, potentially overestimating the clinical impact of wearable sensor technologies. Furthermore, although some studies (3/30, 10%) discussed the potential cost-effectiveness of sensor-based systems, no direct economic evaluations were identified, limiting the ability to substantiate financial feasibility claims.

Technical challenges also persist. Variability in signal-to-noise ratios, sensor displacement errors, and inconsistencies in data collection protocols underscore the need for improved hardware design and standardized preprocessing algorithms. Differences in feature extraction and model architectures limit cross-study comparisons and reproducibility.

Finally, the field urgently requires greater standardization. Variability in sensor placement, protocols, and data interpretation hinders reproducibility and large-scale comparison. Establishing consensus-driven guidelines for wearable sensor research—including standardized task protocols, data reporting frameworks, and model transparency criteria—will be essential to enable scalability, reproducibility, and eventual clinical translation.

Future Directions

To fully realize the potential of wearable sensors in health care, future research should focus on several key areas. Expanding RCTs is essential to establish causal relationships between wearable sensor use and health outcomes beyond feasibility studies. Standardized data analysis frameworks will improve comparability and reproducibility, enabling integration into multicenter trials and large-scale studies. Long-term, multicenter studies will enhance real-world validation and assess sensor accuracy, usability, and adoption across health care settings.

Integration with cloud-based platforms and telemedicine will enhance scalability and enable real-time remote monitoring across diverse populations [42]. Cost-effectiveness analyses will determine financial feasibility and accessibility, supporting broader health care adoption and effective use in resource-limited settings [43].

Conclusions

This systematic review highlights the growing clinical relevance of wearable sensors for rehabilitation monitoring, disease risk assessment, and personalized health care. IMU-based smartwatches, multisensor systems, and gait-monitoring devices demonstrate high accuracy in mobility assessment, fall risk prediction, and chronic disease management for digital health and precision medicine.

Despite their utility, the following challenges remain: small sample sizes, real-world validation gaps, and inconsistent ML methodologies. Future research should standardize protocols, expand clinical trials, and integrate sensors into telemedicine and cloud-based analytics platforms.

Overcoming these challenges will enable wearable sensors to revolutionize health care through real-time, noninvasive monitoring that bridges traditional clinical assessments and continuous real-world health tracking.

Acknowledgments

This research was supported by a grant from the Korean Advanced Research Projects Agency for Health initiative through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (grant RS-2024-00507256). This research was also supported by a grant from the Patient-Centered Clinical Research Coordinating Center funded by the Ministry of Health and Welfare, Republic of Korea (grant RS-2025-02213601).

Data Availability

This systematic review paper does not include original data or materials as it primarily synthesizes and analyzes existing studies. Therefore, there are no specific datasets or materials available for sharing.

Authors' Contributions

BG contributed to the conception and design of the study, data analysis, interpretation of the results, and drafting of the manuscript. Yoo JI supervised the study and served as the corresponding author. All other authors critically reviewed the manuscript and approved the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[PDF File (Adobe PDF File), 103 KB - mhealth_v14i1e76084_app1.pdf]

References

1. Wang X, Yu H, Kold S, Rahbek O, Bai S. Wearable sensors for activity monitoring and motion control: a review. Biomimetic Intell Robot 2023 Mar;3(1):100089. [doi: 10.1016/j.birob.2023.100089]
2. Wei S, Wu Z. The application of wearable sensors and machine learning algorithms in rehabilitation training: a systematic review. Sensors (Basel) 2023 Sep 05;23(18):7667 [FREE Full text] [doi: 10.3390/s23187667] [Medline: 37765724]
3. Prisco G, Pirozzi MA, Santone A, Esposito F, Cesarelli M, Amato F, et al. Validity of wearable inertial sensors for gait analysis: a systematic review. Diagnostics (Basel) 2024 Dec 27;15(1):36 [FREE Full text] [doi: 10.3390/diagnostics15010036] [Medline: 39795564]
4. Liu L, Pu Y, Fan J, Yan Y, Liu W, Luo K, et al. Wearable sensors, data processing, and artificial intelligence in pregnancy monitoring: a review. Sensors (Basel) 2024 Oct 04;24(19):6426 [FREE Full text] [doi: 10.3390/s24196426] [Medline: 39409471]
5. Ortiz BL, Gupta V, Kumar R, Jalin A, Cao X, Ziegenbein C, et al. Data preprocessing techniques for AI and machine learning readiness: scoping review of wearable sensor data in cancer care. JMIR Mhealth Uhealth 2024 Sep 27;12:e59587 [FREE Full text] [doi: 10.2196/59587] [Medline: 38626290]
6. Bolam SM, Batinica B, Yeung TC, Weaver S, Cantamessa A, Vanderboor TC, et al. Remote patient monitoring with wearable sensors following knee arthroplasty. Sensors (Basel) 2021 Jul 29;21(15):5143 [FREE Full text] [doi: 10.3390/s21155143] [Medline: 34372377]
7. Angelucci A, Aliverti A. An IMU-based wearable system for respiratory rate estimation in static and dynamic conditions. Cardiovasc Eng Technol 2023 Jun;14(3):351-363 [FREE Full text] [doi: 10.1007/s13239-023-00657-3] [Medline: 36849621]
8. Greene BR, Premoli I, McManus K, McGrath D, Caulfield B. Predicting fall counts using wearable sensors: a novel digital biomarker for Parkinson's disease. Sensors (Basel) 2021 Dec 22;22(1):54 [FREE Full text] [doi: 10.3390/s22010054] [Medline: 35009599]
9. Gordon MF, Grachev ID, Mazeh I, Dolan Y, Reilmann R, Loupe PS, et al. Quantification of motor function in huntington disease patients using wearable sensor devices. Digit Biomark 2019 Sep 06;3(3):103-115 [FREE Full text] [doi: 10.1159/000502136] [Medline: 32095771]
10. Presley BM, Sklar JC, Hazelwood SJ, Berg-Johansen B, Klisch SM. Balance assessment using a smartwatch inertial measurement unit with principal component analysis for anatomical calibration. Sensors (Basel) 2023 May 09;23(10):4585 [FREE Full text] [doi: 10.3390/s23104585] [Medline: 37430500]
11. De Cannière H, Corradi F, Smeets CJ, Schoutteten M, Varon C, Van Hoof C, et al. Wearable monitoring and interpretable machine learning can objectively track progression in patients during cardiac rehabilitation. Sensors (Basel) 2020 Jun 26;20(12):3601 [FREE Full text] [doi: 10.3390/s20123601] [Medline: 32604829]
12. Nunes AS, Potter İ, Mishra RK, Casado J, Dana N, Geronimo A, et al. Using wearable sensors and machine learning to assess upper limb function in Huntington's disease. Research Square :1-26 Preprint posted online on June 3, 2024 [FREE Full text] [doi: 10.21203/rs.3.rs-4355136/v1]
13. Mahadevan N, Demanuele C, Zhang H, Volfson D, Ho B, Erb MK, et al. Development of digital biomarkers for resting tremor and bradykinesia using a wrist-worn wearable device. NPJ Digit Med 2020 Jan 15;3:5 [FREE Full text] [doi: 10.1038/s41746-019-0217-7] [Medline: 31970290]
14. Seo NJ, Coupland K, Finetto C, Scronce G. Wearable sensor to monitor quality of upper limb task practice for stroke survivors at home. Sensors (Basel) 2024 Jan 16;24(2):554 [FREE Full text] [doi: 10.3390/s24020554] [Medline: 38257646]

15. Odhiambo CO, Ablonczy L, Wright PJ, Corbett CF, Reichardt S, Valafar H. Detecting medication-taking gestures using machine learning and accelerometer data collected via smartwatch technology: instrument validation study. JMIR Hum Factors 2023 May 04;10:e42714 [FREE Full text] [doi: 10.2196/42714] [Medline: 37140971]
16. Hwang TH, Effenberg AO. Head trajectory diagrams for gait symmetry analysis using a single head-worn IMU. Sensors (Basel) 2021 Oct 05;21(19):6621 [FREE Full text] [doi: 10.3390/s21196621] [Medline: 34640945]
17. Wu J, Kuruvithadam K, Schaer A, Stoneham R, Chatzipirpiridis G, Easthope CA, et al. An intelligent in-shoe system for gait monitoring and analysis with optimized sampling and real-time visualization capabilities. Sensors (Basel) 2021 Apr 19;21(8):2869 [FREE Full text] [doi: 10.3390/s21082869] [Medline: 33921846]
18. John J, Soangra R. Visualization-driven time-series extraction from wearable systems can facilitate differentiation of passive ADL characteristics among stroke and healthy older adults. Sensors (Basel) 2022 Jan 13;22(2):598 [FREE Full text] [doi: 10.3390/s22020598] [Medline: 35062557]
19. Meyer BM, Depetrillo P, Franco J, Donahue N, Fox SR, O'Leary A, et al. How much data is enough? A reliable methodology to examine long-term wearable data acquisition in gait and postural sway. Sensors (Basel) 2022 Sep 15;22(18):6982 [FREE Full text] [doi: 10.3390/s22186982] [Medline: 36146348]
20. Toumieux P, Chevalier L, Sahuguède S, Julien-Vergonjanne A. Optical wireless connected objects for healthcare. Healthc Technol Lett 2015 Sep 24;2(5):118-122 [FREE Full text] [doi: 10.1049/htl.2015.0028] [Medline: 26609417]
21. Elstub LJ, Nurse CA, Grohowski LM, Volgyesi P, Wolf DN, Zelik KE. Tibial bone forces can be monitored using shoe-worn wearable sensors during running. J Sports Sci 2022 Aug;40(15):1741-1749 [FREE Full text] [doi: 10.1080/02640414.2022.2107816] [Medline: 35938189]
22. Perraudin CG, Illiano VP, Calvo F, O'Hare E, Donnelly SC, Mullan RH, et al. Observational study of a wearable sensor and smartphone application supporting unsupervised exercises to assess pain and stiffness. Digit Biomark 2018 Oct 23;2(3):106-125 [FREE Full text] [doi: 10.1159/000493277] [Medline: 32095762]
23. Giggins OM, Vavasour G, Doyle J. Unsupervised assessment of frailty status using wearable sensors: a feasibility study among community-dwelling older adults. Adv Rehabil Sci Pract 2025 Feb 15;14:27536351241311845. [doi: 10.1177/27536351241311845] [Medline: 39958411]
24. Savi D, Graziano L, Giordani B, Schiavetto S, Vito CD, Migliara G, et al. New strategies of physical activity assessment in cystic fibrosis: a pilot study. BMC Pulm Med 2020 Oct 30;20(1):285 [FREE Full text] [doi: 10.1186/s12890-020-01313-5] [Medline: 33126875]
25. Haghi M, Ershadi A, Deserno TM. Recognizing human activity of daily living using a flexible wearable for 3D spine pose tracking. Sensors (Basel) 2023 Feb 12;23(4):2066 [FREE Full text] [doi: 10.3390/s23042066] [Medline: 36850664]
26. 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 [FREE Full text] [doi: 10.1159/000488040] [Medline: 32095756]
27. Sun R, Aldunate RG, Sosnoff JJ. The validity of a mixed reality-based automated functional mobility assessment. Sensors (Basel) 2019 May 11;19(9):2183 [FREE Full text] [doi: 10.3390/s19092183] [Medline: 31083514]
28. Ramezani R, Zhang W, Xie Z, Shen J, Elashoff D, Roberts P, et al. A combination of indoor localization and wearable sensor-based physical activity recognition to assess older patients undergoing subacute rehabilitation: baseline study results. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e14090 [FREE Full text] [doi: 10.2196/14090] [Medline: 31293244]
29. Liew BX, Crisafulli O, Evans DW. Quantifying lumbar sagittal plane kinematics using a wrist-worn inertial measurement unit. Front Sports Act Living 2024 May 14;6:1381020 [FREE Full text] [doi: 10.3389/fspor.2024.1381020] [Medline: 38807615]
30. Kwon S, Honegger K, Mason M. Daily physical activity among toddlers: hip and wrist accelerometer assessments. Int J Environ Res Public Health 2019 Nov 01;16(21):4244 [FREE Full text] [doi: 10.3390/ijerph16214244] [Medline: 31683776]
31. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. mActive: a randomized clinical trial of an automated mhealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):e002239 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
32. Hup RG, Linssen EC, Eversdijk M, Verbruggen B, Bak MA, Habibovic M, et al. Rationale and design of the BECA project: smartwatch-based activation of the chain of survival for out-of-hospital cardiac arrest. Resusc Plus 2024 Feb 09;17:100576 [FREE Full text] [doi: 10.1016/j.resplu.2024.100576] [Medline: 38370313]
33. Browne S, Kechadi MT, O'Donnell S, Dow M, Tully L, Doyle G, et al. Mobile health apps in pediatric obesity treatment: process outcomes from a feasibility study of a multicomponent intervention. JMIR Mhealth Uhealth 2020 Jul 08;8(7):e16925 [FREE Full text] [doi: 10.2196/16925] [Medline: 32673267]
34. Burns D, Razmjou H, Shaw J, Richards R, McLachlin S, Hardisty M, et al. Adherence tracking with smart watches for shoulder physiotherapy in rotator cuff pathology: protocol for a longitudinal cohort study. JMIR Res Protoc 2020 Jul 05;9(7):e17841 [FREE Full text] [doi: 10.2196/17841] [Medline: 32623366]
35. Bailey DP, Ahmed I, Cooper DL, Finlay KA, Froome HM, Nightingale TE, et al. Validity of a wrist-worn consumer-grade wearable for estimating energy expenditure, sedentary behaviour, and physical activity in manual wheelchair users with spinal cord injury. Disabil Rehabil Assist Technol 2025 Apr;20(3):708-714 [FREE Full text] [doi: 10.1080/17483107.2024.2405895] [Medline: 39301994]

36. Porciuncula F, Roto AV, Kumar D, Davis I, Roy S, Walsh CJ, et al. Wearable movement sensors for rehabilitation: a focused review of technological and clinical advances. PM R 2018 Sep;10(9 Suppl 2):S220-S232 [FREE Full text] [doi: 10.1016/j.pmrj.2018.06.013] [Medline: 30269807]
37. Merritt MW. Health researchers' ancillary care obligations in low-resource settings: how can we tell what is morally required? Kennedy Inst Ethics J 2011 Dec;21(4):311-347 [FREE Full text] [doi: 10.1353/ken.2011.0019] [Medline: 22187929]
38. Tao Q, Liu S, Zhang J, Jiang J, Jin Z, Huang Y, et al. Clinical applications of smart wearable sensors. iScience 2023 Jul 27;26(9):107485 [FREE Full text] [doi: 10.1016/j.isci.2023.107485] [Medline: 37636055]
39. Tranquilli C, Bernetti A, Picerno P. Ambulatory joint mobility and muscle strength assessment during rehabilitation using a single wearable inertial sensor. Medicina dello Sport 2013;66(4):583-597 [FREE Full text]
40. Ajčević M, Deodato M, Murena L, Miladinovic A, Mezzarobba S, Accardo A. Assessment of mobility deficit and treatment efficacy in adhesive capsulitis by measurement of kinematic parameters using IMU sensors. In: Proceedings of the IEEE International Symposium on Medical Measurements and Applications. 2020 Jul 10 Presented at: MeMeA '20; June 1-July 1, 2020; Bari, Italy URL: https://ieeexplore.ieee.org/document/9137157 [doi: 10.1109/MeMeA49120.2020.9137157]
41. Parel I, Candoli V, Filippi MV, Padolino A, Merolla G, Sanniti S, et al. Shoulder rehabilitation exercises with kinematic biofeedback after arthroscopic rotator cuff repair: protocol for a new integrated rehabilitation program. JMIR Res Protoc 2023 Mar 22;12:e35757 [FREE Full text] [doi: 10.2196/35757] [Medline: 36947146]
42. Junaid SB, Imam AA, Balogun AO, De Silva LC, Surakat YA, Kumar G, et al. Recent advancements in emerging technologies for healthcare management systems: a survey. Healthcare (Basel) 2022 Oct 03;10(10):e39 [FREE Full text] [doi: 10.3390/healthcare10101940] [Medline: 36292387]
43. De Sario Velasquez GD, Borna S, Maniaci MJ, Coffey JD, Haider CR, Demaerschalk BM, et al. Economic Perspective of the Use of Wearables in Health Care: A Systematic Review. Mayo Clin Proc Digit Health 2024 Sep;2(3):299-317 [FREE Full text] [doi: 10.1016/j.mcpdig.2024.05.003] [Medline: 40206120]

Abbreviations

AUROC: area under the receiver operating characteristic curve
IMU: inertial measurement unit
JBI: Joanna Briggs Institute
ML: machine learning
MS: multiple sclerosis
NOS: Newcastle-Ottawa Scale
PCA: principal component analysis
PD: Parkinson disease
POMA: Performance-Oriented Mobility Assessment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: version 2 of the Cochrane risk-of-bias tool for randomized trials
SVM: support vector machine
TUG: Timed Up and Go

Edited by M Giurgiu; submitted 17.Apr.2025; peer-reviewed by A Uchenna, P Picerno; comments to author 26.Jul.2025; accepted 17.Nov.2025; published 09.Jan.2026.

Please cite as:
Gu B, Kim HS, Kim H, Yoo JI
Advancements in Wearable Sensor Technologies for Health Monitoring in Terms of Clinical Applications, Rehabilitation, and Disease Risk Assessment: Systematic Review
JMIR Mhealth Uhealth 2026;14:e76084
URL: https://mhealth.jmir.org/2026/1/e76084
doi:10.2196/76084
PMID:

Review

Evaluating the Effectiveness of Digital Interventions for Stress Management in Pregnant Women: Systematic Review and Meta-Analysis

Jeung-Im Kim[1], PhD; Joo Yun Lee[2], PhD; So-Hee Park[3], MSN

[1]School of Nursing, College of Medicine, Soonchunhynag University, Asan-si, Republic of Korea

[2]College of Nursing, Gachon University, Incheon, Republic of Korea

[3]School of Nursing, College of Medicine, Soonchunhyang University, Asan-si, Republic of Korea

Corresponding Author:
Joo Yun Lee, PhD
College of Nursing, Gachon University
191, Hambangmoe-ro, Yeonsu-gu
Incheon, 21936
Republic of Korea
Phone: 82 10 9169 0130
Email: jooyun@gachon.ac.kr

Abstract

Background: Psychological stress during pregnancy is common and has been associated with adverse maternal and neonatal outcomes. Digital health interventions (DHIs) have emerged as a scalable approach to support stress management during pregnancy, yet evidence remains fragmented, and prior reviews have largely focused on broad perinatal mental health outcomes or delivery platforms rather than stress-specific effects and targeted intervention components.

Objective: This systematic review and meta-analysis aimed to evaluate the effectiveness of DHIs specifically designed to reduce stress during pregnancy and to examine how intervention strategies and delivery methods are associated with stress outcomes.

Methods: We conducted a systematic review and meta-analysis following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines. Randomized controlled trials and quasi-experimental studies involving pregnant women were eligible if they evaluated any digitally delivered intervention—such as mobile apps, web-based programs, or telemedicine—intended to reduce stress, and reported validated stress outcomes. We searched CINAHL, the Cochrane Library, Embase, and PubMed from database inception through November 2025. Risk of bias was assessed using the Cochrane risk of bias 2 tool for randomized trials and the risk of bias in nonrandomized studies of interventions tool for nonrandomized studies. Where appropriate, effect sizes were pooled using random-effects meta-analysis with the Hartung–Knapp–Sidik–Jonkman method and reported as standardized mean differences.

Results: A total of 19 studies were included. Overall, DHIs were associated with a significant reduction in stress compared with control conditions (standardized mean difference –0.45, 95% CI –0.59 to –0.32; 95% prediction interval –0.78 to –0.13), with low to moderate heterogeneity. Strategy-based subgroup analyses indicated that mindfulness- and education-focused interventions showed favorable effects, but formal tests for between-subgroup differences were not statistically significant. Evidence certainty was rated as moderate, primarily due to risk-of-bias concerns in some trials.

Conclusions: This review provides stress-focused evidence that DHIs can support stress reduction during pregnancy and extends existing literature by systematically disaggregating interventions according to delivery methods, functional features, and content strategies. This study offers a component-oriented synthesis that informs the design and selection of digital stress-management interventions for pregnant women. In real-world antenatal care, these tools may complement clinician-delivered services by expanding access to low-intensity, scalable support, particularly when interventions integrate skills-based content with supportive digital functions. Future research should directly compare single versus combined strategies and evaluate implementation across diverse populations and care settings.

(JMIR Mhealth Uhealth 2026;14:e66267) doi:10.2196/66267

KEYWORDS

pregnancy; stress management; digital health; mHealth; telemedicine; systematic review; meta-analysis

Introduction

Overview

Pregnancy represents a major life transition that is frequently accompanied by heightened psychological stress, including pregnancy-specific stress related to concerns about maternal health, fetal well-being, and childbirth. Recent evidence indicates that pregnancy-specific stress is prevalent even among low-risk populations and tends to increase with perceived pregnancy risk and advancing gestational age [1]. Elevated stress during pregnancy has been consistently associated with adverse maternal mental health outcomes as well as negative neonatal and child health outcomes, underscoring the importance of timely and effective stress management during the antenatal period [2-4].

A range of nonpharmacological interventions has been developed to address stress during pregnancy, and their effectiveness has been demonstrated across multiple studies. These include psychoeducational programs, structured psychotherapies such as cognitive behavioral therapy and interpersonal psychotherapy, and mind–body approaches including mindfulness and yoga [5-8]. Traditionally, such interventions have been delivered through face-to-face sessions. While effective, in-person delivery can be constrained by barriers related to access, cost, time, geographic distance, and the availability of trained professionals—limitations that may be particularly salient during pregnancy.

Digital health interventions (DHIs) have emerged as a scalable and low-threshold approach for delivering stress-management strategies, with growing evidence supporting their effectiveness in the general population. Meta-analytic evidence indicates that app-based stress-management interventions yield small but statistically significant improvements across self-reported, physiological, and neuroendocrine stress outcomes [9], and similar modest reductions in perceived stress have been reported in randomized trials of mental health smartphone apps [10]. These findings suggest that digitally delivered interventions can effectively support stress reduction when designed and implemented appropriately.

However, evidence specific to pregnancy remains more limited and less conclusive. Much of the existing literature on DHIs during pregnancy has focused on broader perinatal mental health outcomes—particularly depression and anxiety—rather than stress as a primary target. For example, reviews of digital mindfulness and nurse-led eHealth interventions have reported consistent benefits for depressive and anxiety symptoms but mixed or inconclusive effects on stress [11,12], while telemedicine-based psychological interventions often combine antenatal and postpartum populations and rarely prioritize stress outcomes [13]. An umbrella review of DHIs for perinatal women suggested overall benefits for psychological outcomes, including stress; however, its findings were pooled across heterogeneous populations, intervention types, outcomes (stress, depression, and anxiety), and perinatal stages, with only a small subset of reviews specifically addressing stress during pregnancy [14]. Collectively, these syntheses highlight that stress-focused evidence in pregnant populations remains fragmented and underdeveloped.

At the same time, DHIs offer distinct methodological and practical advantages that warrant more nuanced evaluation. Digital interventions are inherently multicomponent, comprising combinations of delivery methods (eg, mobile apps, web-based platforms, and telemedicine), functional features (eg, self-monitoring, automated feedback, and reminders), and intervention contents (eg, psychoeducation, mindfulness, relaxation, and cognitive behavioral skills). Behavioral science frameworks emphasize that technology functions primarily as a delivery vehicle rather than the therapeutic agent itself [15], and empirical assessments of stress-management apps similarly distinguish delivery modalities, functional features, and therapeutic content as separable but interacting components [16]. Despite this, prior reviews of digital mental health interventions during pregnancy have rarely synthesized evidence across these dimensions, limiting understanding of which combinations of delivery methods, functions, and contents are most relevant for stress reduction in pregnant women.

Rationale

Taken together, the existing literature highlights a persistent gap in evidence regarding digital interventions specifically designed to reduce stress during pregnancy. Although DHIs for perinatal mental health have received growing attention, most prior reviews have focused primarily on depression and anxiety or have pooled multiple psychological outcomes, with stress often insufficiently examined as a primary outcome. As a result, the effectiveness of digital interventions targeting stress during pregnancy remains unclear.

Moreover, previous syntheses have largely categorized digital interventions by delivery platform, providing limited insight into how specific intervention components contribute to stress reduction. DHIs are inherently multicomponent, integrating delivery methods, functional features, and therapeutic content; however, the relative importance of these components for managing pregnancy-related stress has not been systematically examined.

To address these gaps, the present systematic review and meta-analysis provide an updated and focused synthesis of digital stress-management interventions during pregnancy. By centering stress reduction as the primary outcome and examining intervention effectiveness through the lens of delivery methods and intervention content, this review aims to clarify how digital interventions can be optimally designed and implemented to support stress management in real-world antenatal care settings.

Objectives

Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidance [17] and PRISMA-S [18], this systematic review aimed to (1) identify and describe digital stress-management interventions for

pregnant women (Population), including delivery platforms and key components (Intervention); (2) evaluate the effectiveness of these interventions compared with usual care, wait-list, or active comparators (Comparator) on validated stress outcomes (Outcome); and (3) summarize intervention strategies and delivery methods associated with engagement and stress reduction.

Methods

Overview

We conducted a systematic review and meta-analysis to evaluate digitally delivered interventions for reducing stress in pregnant women. The review adhered to the PRISMA guidelines [17] and PRISMA-S (PRISMA literature search extension) [18] (Multimedia Appendix 1).

Eligibility Criteria

Eligibility criteria were predefined using the PICOS (population, intervention, comparison, outcome, and study design) framework. Participants were pregnant women. Interventions included any digitally delivered program primarily aimed at reducing or managing stress during pregnancy (antenatal period), delivered via mobile apps, web-based platforms, telemedicine, or other technology-enabled modalities. Studies in which the intervention was initiated postpartum were excluded. Comparators included usual care, waitlist/no intervention, or nondigital interventions. Outcomes included stress assessed using validated instruments; when the Depression Anxiety Stress Scales (DASS) was used, only the stress subscale was extracted. Study designs included RCTs and quasi-experimental/nonrandomized intervention studies with extractable pre-post outcome data.

Exclusion criteria included (1) postpartum-only interventions (interventions initiated after delivery), (2) interventions not primarily targeting stress management (eg, programs targeting depression, weight control, smoking cessation, or other conditions in which stress was only a secondary outcome without a stress-focused rationale), (3) insufficient outcome data to estimate effects (no extractable pre-post data or between-group comparisons), and (4) conference abstracts or trial registry entries without sufficient methodological detail or outcome data for extraction and risk-of-bias assessment.

Information Sources

The initial systematic search was conducted across 4 databases (CINAHL, Cochrane Library, Embase, and PubMed) from database inception to September 2023. In response to reviewer comments and to ensure the currency of the evidence base, the search was updated by re-running the same search strategy in all databases for records published from September 2023 through November 30, 2025. The updated search was executed on December 10, 2025, by JYL and SHP, and all newly retrieved records were screened and incorporated into the final synthesis as appropriate. Searches were conducted with the assistance of 2 medical librarians, and the final search strategies for each database are reported in Multimedia Appendix 2. Each database was searched independently using its native platform (PubMed, Embase.com, CINAHL via EBSCOhost, and the Cochrane Library). Databases were not searched simultaneously on a single platform. In addition to database searching, reference lists of included studies were manually screened to identify potentially relevant articles using citation searching methods. We did not contact study authors, experts, or intervention developers to obtain additional data, nor were other information sources such as trial registries or gray literature repositories searched.

Search Strategy

Electronic search strategies were developed in consultation with 2 medical librarians, who proposed candidate keywords and controlled vocabulary; the research team reviewed and refined the strategies, and final searches were executed by the study team with librarian support. All search strategies were newly developed for this review and were not adapted from or reused based on search strategies reported in previous literature reviews. The strategy combined controlled vocabulary (MeSH [Medical Subject Headings], Emtree, and CINAHL Subject Headings) and keywords related to pregnancy, digital health, and stress. Pregnancy-related terms included “pregnan*,” “pregnancy,” “antenatal,” and “prenatal” (and related indexing terms). Digital health terms included “digital health,” “mHealth,” “eHealth,” “telemedicine,” “mobile app,” and “smartphone.” Stress-related terms included “stress” and relevant intervention terms (eg, relaxation, mindfulness, cognitive behavioral). Animal studies were excluded, and results were limited to English-language publications. Full search strategies for each database are provided in Multimedia Appendix 2.

Selection Process

All records were imported into EndNote and deduplicated. Three reviewers (JIK, JYL, and SHP) independently screened titles and abstracts to identify potentially eligible studies. Full texts were retrieved for records deemed relevant or uncertain; when full-text articles were difficult to obtain, a medical librarian assisted with document retrieval. Two reviewers (JIK and SHP) independently assessed full-text eligibility and cross-checked decisions, with disagreements resolved through discussion (and adjudication by a third reviewer when necessary). For the updated search, a second screening pass was completed on December 12, 2025, and all newly retrieved articles were screened and assessed for eligibility using the same procedures as the initial review, with eligible studies incorporated into the final synthesis.

Data Collection Process and Data Items

Data was extracted using a piloted structured form capturing intervention purpose, participant characteristics, timing and duration, intervention strategies/components and digital functions, delivery mode, comparator, sample size, stress measures, and outcome values. When outcomes were measured repeatedly, the assessment closest to intervention completion was extracted; follow-up outcomes were extracted separately where available. Two reviewers independently entered data into an Excel file and verified consistency. Discrepancies were resolved by consensus among 3 reviewers to ensure accuracy and consistent terminology.

Study Risk of Bias Assessment

Risk of bias was assessed using RoB 2 (a revised Cochrane risk-of-bias tool for randomized trials) for randomized trials [19] and ROBINS-I (risk of bias in nonrandomized studies of interventions) for quasi-experimental/nonrandomized studies [20]. Two reviewers (JIK and SHP) independently rated each study, and disagreements were resolved through discussion with adjudication by a third reviewer [JYL] when needed. RoB 2 judgments were made across 5 domains and summarized as low risk, some concerns, or high risk. ROBINS-I judgments were made across 7 domains and summarized as low, moderate, serious, or critical risk of bias. Risk-of-bias assessments informed sensitivity analyses and the certainty of evidence (GRADE [Grading of Recommendations, Assessment, Development and Evaluation]).

Effect Measures

The primary outcome was the change in stress from baseline to postintervention (closest to intervention completion). When DASS was used, only the stress subscale was extracted. Continuous outcomes were synthesized as standardized mean differences (SMDs; Hedges *g*) and reported with 95% CIs [21], with negative values indicating lower stress in the intervention group. When studies did not report means and SDs directly, effect estimates were derived from available statistics (eg, *F* values, odds ratios, chi-square values) using standard conversion methods.

Synthesis Methods

Study characteristics (population, intervention components and digital functions, delivery mode, stress measures, and effects) were summarized narratively. For quantitative synthesis, meta-analyses were conducted using random-effects models to reflect expected between-study variability in true effects. Analyses were performed in in R (version 4.5.2; R Foundation for Statistical Computing, Vienna, Austria) using the Hartung–Knapp–Sidik–Jonkman method [22]. Heterogeneity was quantified using τ^2 and I^2, with the Cochran Q test reported where appropriate. To distinguish the average pooled effect from the expected distribution of effects across settings, we reported 95% prediction intervals for the main meta-analyses (but not for subgroup analyses). CIs describe uncertainty around the pooled mean effect, whereas prediction intervals indicate the plausible range of effects in a new setting. Prespecified subgroup analyses (eg, intervention strategies/components and delivery modes) and, when sufficient studies were available, meta-regression were used to explore potential sources of heterogeneity. Sensitivity analyses excluded studies at high or serious risk of bias and tested alternative assumptions for imputed parameters where applicable. Small-study effects were assessed using funnel plots and Egger's regression test when ≥10 studies were available for a given main analysis.

Certainty Assessment

Two reviewers (JIK and SHP) independently assessed certainty of evidence using GRADE, with disagreements resolved through consensus and third-reviewer adjudication. Certainty was evaluated across risk of bias, inconsistency, indirectness, imprecision, and publication bias. Downgrading decisions were documented in the Summary of Findings table and were informed by RoB 2/ROBINS-I judgments, heterogeneity (including the extent to which prediction intervals indicated variability across settings), imprecision (width of CIs and information size), and evidence of small-study effects (funnel/Egger).

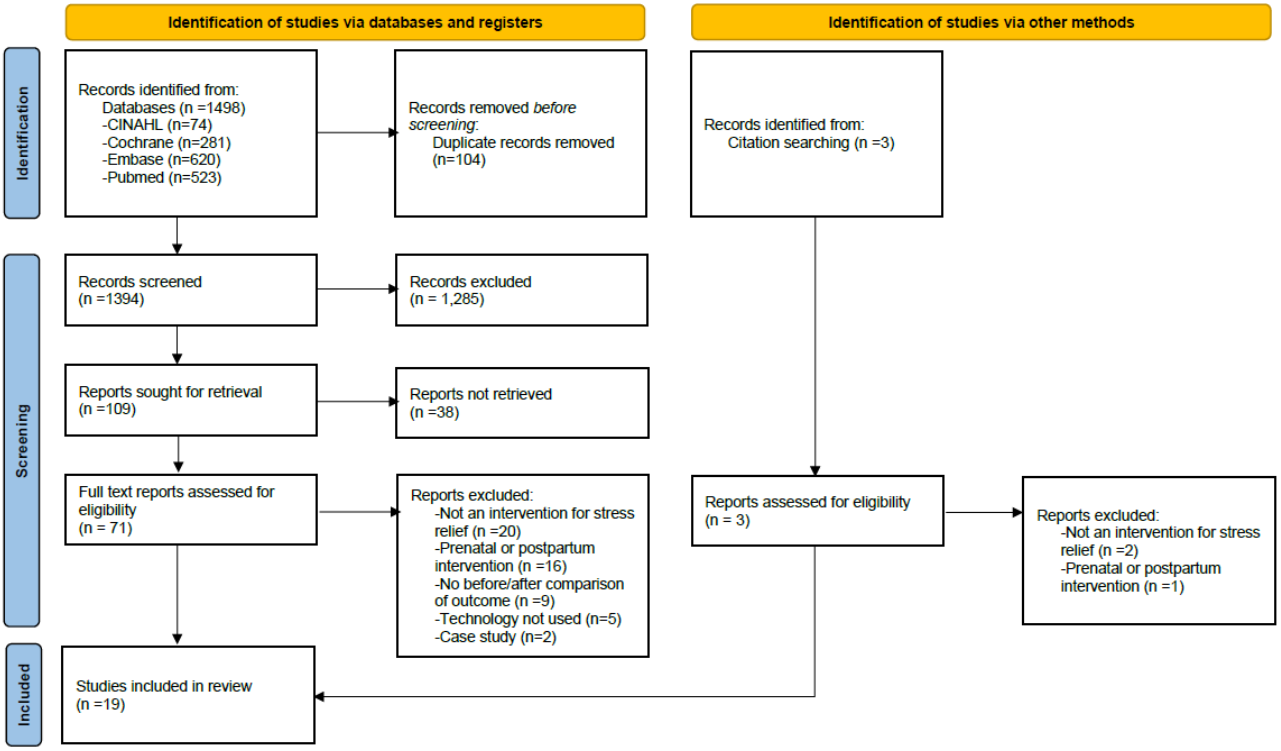
Results

Study Selection

The study selection process is summarized in Figure 1. Studies were excluded if they targeted health conditions other than stress (eg, depression or anxiety, weight control, or smoking cessation) without a stress-focused intervention rationale, or if stress was measured only as a secondary outcome. Studies describing intervention development without reporting evaluable effectiveness outcomes were also excluded. Citation searching identified additional records, which were assessed at the full-text level but ultimately did not meet the inclusion criteria. Reasons for exclusion at each stage were documented and are presented in the PRISMA flow diagram.

Following the updated search and eligibility assessment of newly retrieved records, 4 additional studies were included. In total, 19 studies (15 from the initial search and 4 from the updated search) were included in the final review.

Figure 1. Study selection process based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flow diagram.



Study Characteristics

Included studies were published between 2016 and 2025 [23-41]. Among them, 7 studies [24,26,28,31,33,34,38] were conducted in the United States (37%), followed by studies from Iran (n=2) [29,37], China (n=2) [25,34], Korea (n=2) [24,32], and other countries. Overall, there were 12 RCTs (63%) and 7 quasi-experimental studies (37%; Table 1).

Table 1. Characteristics of the included studies.

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Randomized controlled study							
Balderas-Díaz et al [23]	Spain (2022)	At risk of having SGA[a] fetuses	Second trimester and part of the third trimester	16 weeks	Mobile	Education and information	• Learn about specific topics, carry out the proposed activities to strengthen training • Four components: Medical advice, health care, communication with the fetus for stimulation purposes, and emotional management
Lee et al [24]	Korea (2023)	Working pregnant women	From ≤34 weeks	4 weeks	Mobile	Education and information with peer support	• Four components: Eight education sessions, a health log, a diary, and an anonymous discussion board • Each session lasts approximately 15 minutes • Four targeted practices: Sleep and rest, eating, physical activity, and stress management
Sun et al [25]	China (2021)	At risk of perinatal depression (with an EPDS[b] score >9 or a PHQ-9[c] score >4)	12-20 weeks	8 weeks	Mobile	Mindfulness	• Formal mindfulness training: Body scan, mindful breathing, mindful stretching, and mindful meditation lasting 15-25 minutes per day • Informal training: Pausing in the midst of daily life, mindful eating, mindful walking, and 3-minute breathing practices
Smith et al [26]	United States (2021)	Obstetric patients of outpatient clinic during the COVID-19 pandemic	14-34 weeks	30 days	Mobile	Mindfulness	• Mindfulness meditation, sleep stories, and nature sounds • Encourage use of 70 minutes per week, preferably 10 minutes per day
Krusche et al [27]	United Kingdom (2018)	Pregnant women recruited online	12-34 weeks	4 weeks	Website	Mindfulness	• Formal and informal meditation practices such as body scan, mindful movement, breathing space, and mindful eating • Online sessions, assignments, and emails
Dennis-Tiwary et al [28]	United States (2017)	Pregnant women recruited from a large urban hospital	19-29 weeks	4 weeks	Mobile	Game	• Gamified attention bias modification training (ABMT), incorporating video game-like features such as animated characters and sound effects • 10 rounds each day (≤10 minutes of play) for 4 days/week: 160 rounds total for the duration of the study
Kia [29]	Iran (2023)	No stressful event other than COVID-19 disease in the preceding 6 months	12 or more weeks	3 weeks	Mobile	Education and information	• Educational theme: (1) COVID-19 disease explanation, personal hygiene, and transmission modes, (2) COVID-19 in pregnancy and childbirth, (3) COVID-19 and breastfeeding, (4) COVID-19 and infants • Five sessions of 30 minutes, 2 sessions per week

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Mauriello et al [30]	United States (2016)	Pregnant women recruited from 6 locations of 3 federally funded health centers	From <19 weeks	24 weeks	Mobile	Education and information	• Stage-matched and tailored guidance, 3 interactive sessions focused on 2 priority health behavior risks (smoking, stress management, fruits and vegetables), individually tailored and stage-matched change strategies • Feedback messages within stage-matched activities, including tools such as calculators, quizzes, action plans, support messages, and recipe ideas • Three antenatal sessions are spaced approximately 12 weeks apart, with 2 postpartum assessment-only sessions at 1 month and 4 months
Chua et al [31]	Singapore (2024)	Heterosexual couples recruited from a public tertiary hospital	From >24 weeks	From 24 weeks to 1 month postpartum	Mobile	Education and information and support	• Six components: (1) Education center (multimedia resources, chatbot), (2) Ask an expert (Experienced nurses/midwives will respond to queries within 24 h), (3) Smile center (mood rating, guided mindfulness), (4) Positivity space (support forum), (5) Gamification features (virtual badges, rewards), (6) Helpline
Park et al [32]	Korea (2025)	Pregnant women recruited from an obstetrics and gynecology center	1-32 weeks	8 weeks	Mobile	Mindfulness	• Four sections: (1) breathing mindfulness meditation, (2) body scanning, (3) emotional awareness, (4) self-kindness mindfulness • Each session was divided into 2 subsessions—instruction and practice sessions. • Instructed to practice each session at least twice
Tandon et al [33]	United States (2025)	Pregnant women recruited from 6 university affiliated prenatal care clinics	From <22 weeks	maximum of 14 weeks	Mobile, in-person	Education and Mindfulness	• Twelve-session manualized intervention: CBT[d] content related to behavioral activation, identification, and reframing unhelpful thought patterns, and promotion of positive interactions • Includes various mindfulness practices • Received just-in-time (JIT) messages were sent within 24 h of an elevated stress reading
Tian et al [34]	China (2025)	Couples expecting their first child recruited from outpatient centers	12-20 weeks	6 weeks	Mobile	Mindfulness	• Six one-week modules, each with a thematic session and 6 audio-guided home practice sessions • Mindfulness exercises that involved couple collaboration and interaction
Quasi-experimental study							

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Porter et al [35]	United States (2022)	Pregnant women recruited from a university hospital	<15 weeks to 28 weeks	13 weeks	Mobile	Mindfulness	• Meditations tailored to the trimester and specific physical and emotional states of participants • Recommend daily usage of 10-20 minutes
Jallo et al [36]	United States (2017)	Hospitalized with preterm labor	22+0-33+6 weeks	8 days	Mobile	Education and relaxation	• Educational overview (stress, stress response, and impact on health) • Four guided imagery audio files, including relaxation, focused breathing, positive affirmations, and multiple multisensory images, each lasting 15-20 minutes long • A stress self-assessment scale
Hashemzahi et al [37]	Iran (2022)	With perceived stress (PSS[e] score ≥21.8) and moderate to severe anxiety (CDAS[f] score ≥17)	20-28 weeks	2 weeks	Telemedicine (WhatsApp)	Education and support	• Educational content covered topics such as familiarity with COVID-19, its effect on pregnancy, prevention and self-care during pregnancy, and guidelines for care and prevention during childbirth, postpartum, and breastfeeding • The content was shared with the group every other day over 2 weeks, comprising 6 sessions via WhatsApp. • The researcher followed up to ensure that participants watched the educational videos and addressed any questions or concerns raised by participants
Tsai et al [38]	Taiwan (2018)	Low-risk pregnant women in the outpatient department of a medical center	From 16-24 to 36-38 weeks	12-22 weeks	Website (accessible via smartphone and computer)	Education and information	• Four modules: Maternity health records, antenatal health education, self-management journals, newborn birth records • Participants could store their own information and records • Automatic pop-up windows provide antenatal health information according to a woman's gestational age
Buultjens et al [39]	Australia (2023)	Low-risk pregnant women at one community health site	From 28-30 weeks to 36-38 weeks	6-10 weeks (Antenatal)	Telemedicine (video conference)	Education and support	• One-to-one pregnancy clinical care, integrating standard pregnancy health assessments with structured online small-group interdisciplinary education and peer support, thus incorporating broader psychosocial aspects • Four group antenatal and 4 group postnatal education sessions
Kubo et al [40]	United States (2021)	With moderate-to-moderately-severe depressive symptoms (PHQ-8[g] score 10-19)	From <28 weeks	6 weeks	Mobile	Mindfulness	• Basic mindfulness condition- or situation-specific courses • Headspace courses including breathing exercises, body scan, noting, and visualization • Additional short (1-2 min) lecture videos designed to increase the understanding of mindfulness and encourage its integration into daily life • Participants asked to use the app for 10-20 min a day over 6 weeks

Author	Country (year)	Target	Gestational age at intervention	Intervention period	Delivery mode	Intervention contents	Intervention details
Barber and Masters-Awatere [41]	New Zealand (2022)	Pregnant women recruited from midwifery clinics, through antenatal educators, and via social media	<24 weeks	12 weeks	Mobile	Education and relaxation	• Four types of modules: (1) interactive self-assessments with associated feedback, (2) activities for relaxation, stress management and planning for parenthood, (3) information about psychological and social changes in pregnancy, and (4) discussions with a partner or support person to build social support and interactive reflection and problem-solving

[a]SGA: small for gestational age.

[b]EPDS: Edinburgh Postnatal Depression Scale.

[c]PHQ: patient health questionnaire.

[d]CBT: cognitive behavioral therapy.

[e]PSS: Perceived Stress Scale.

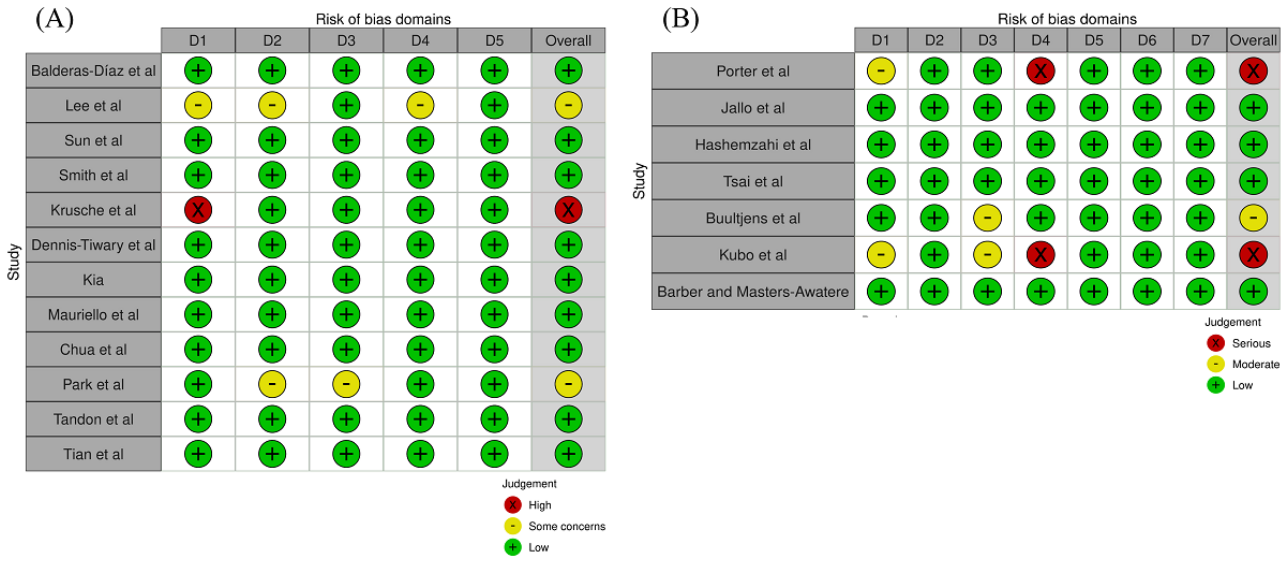
[f]CDAS: Corona Disease Anxiety Scale.

[g]PHQ-8: Patient Health Questionnaire-8.

Risk of Bias in Studies

The results of the risk of bias assessment are presented in Figure 2 [23-41]. Among the 12 RCTs assessed with RoB 2, one study was judged at high risk of bias overall, driven by concerns in the randomization process (Domain 1). Two additional RCTs were rated as having some concerns overall, reflecting incomplete reporting or concerns in specific domains, while the remaining RCTs were judged at low risk of bias across domains.

Figure 2. Risk of bias assessments. (A) Risk of bias for randomized controlled trials assessed using the revised Cochrane risk-of-bias tool. Domains: D1 (randomization process), D2 (deviations from intended interventions), D3 (missing outcome data), D4 (measurement of the outcome), and D5 (selection of the reported result). (B) Risk of bias for nonrandomized studies assessed using the risk of bias in nonrandomized studies of interventions. Domains: D1 (confounding), D2 (selection of participants), D3 (classification of interventions), D4 (deviations from intended interventions), D5 (missing data), D6 (measurement of outcomes), and D7 (selection of the reported result) [23-41].



Among the 7 nonrandomized studies assessed with ROBINS-I, 2 studies were judged at serious risk of bias overall, primarily due to deviations from intended interventions (Domain 4). One study was judged at moderate risk of bias, and the remaining studies were judged at low risk across ROBINS-I domains. No study was judged at critical risk of bias.

Results of Individual Studies

Study Population

All studies targeted pregnant women, with 12 studies focusing on low-risk pregnancies (63%), while the remaining studies targeted pregnant women with specific issues or risk factors. Two studies focused on women at risk of perinatal depression, while one study each targeted women with perceived stress,

stress induced by the COVID-19 pandemic, small for gestational age fetuses, and those hospitalized due to preterm labor. Additionally, one study focused on stress management in working pregnant women. Among the 12 studies, 9 studies in the low-risk group primarily started in the second trimester and often continued into the third trimester. Interventions targeting pregnant women with depression or stress and those carrying small for gestational age fetuses started in the second trimester. Interventions for preterm labor were initiated between 22 and 33 weeks of gestation. Meanwhile, interventions for working pregnant women could be initiated at any time before 34 weeks of gestation.

Intervention Strategies

When categorizing the intervention strategies, 4 studies provided only education and information (21%), while 7 studies incorporated additional strategies (37%). Specifically, 2 studies included professional support, and 2 studies incorporated peer support, which we classified as the strategy of "seeking support." Three studies included relaxation content, such as guided imagery audio files or activities for relaxation, such as watching humorous YouTube videos. One of these studies also included peer support. Furthermore, 7 studies primarily focused on mindfulness, and one study offered an intervention in the form of a game (Table 2).

Table 2. Intervention strategies and functions included in the stress management intervention.

Category and details	Study
Intervention strategies	
Education and information	
Education, medical advice: text, video files, automatic pop-up, or videoconference	Barber and Masters-Awatere [41], Balderas-Díaz et al [23], Lee et al [24], Kia [29], Jallo et al [36], Hashemzahi et al [37], Tsai et al [38], Buultjens et al [39], Chua et al [31], Tandon et al [33]
Seeking support	
Anonymous discussion board	Lee et al [24], Chua et al [31]
Online group peer support	Buultjens et al [39]
Conversation with the partner, other supporters	Barber and Masters-Awatere [41]
Question and answer from the researcher	Hashemzahi et al [37]
Relaxation	
Guided imagery audio files	Barber and Masters-Awatere [41], Jallo et al [36]
Activities for relaxation	Barber and Masters-Awatere [41]
Mindfulness	
Formal training: body scan, mindful breathing, mindful stretching, mindful meditation, and visualization: text, video files, and audio files	Sun et al [25], Smith et al [26], Krusche et al [27], Porter et al [35], Kubo et al [40], Park et al [32], Tandon et al [33], Tian et al [34]
Informal training: mindful eating, mindful walking	Sun et al [25], Krusche et al [27]
Game	
Attention bias modification training	Dennis-Tiwary et al [28]
Functions	
Monitoring (recording)	
Health log, diary	Lee et al [24], Tsai et al [38], Tandon et al [33]
Mindfulness journal	Sun et al [25]
Self-assessment scale	Barber and Masters-Awatere [41], Jallo et al [36]
Automated feedback	
Tailored guidance, feedback messages: programmed messages and chatbot	Barber and Masters-Awatere [41], Mauriello et al [30], Chua et al [31], Tandon et al [33]
Reminder	
Notifications and reminders to prompt use	Barber and Masters-Awatere [41], Sun (WeChat) et al [25], Krusche (e-mail) et al [27], Kubo et al [40], Tian et al [34]

In the process of content extraction, some elements were classified as functions according to the categorization by Paganini et al [16]. For instance, monitoring (recording) included studies where participants regularly assessed stress or fetal development through health logs, diaries, or self-assessment tools. Automated feedback involved the delivery of tailored guidance and feedback messages through programmed communication. There was an intervention incorporated with

interactive chatbot functions that responded to users' questions in real time, providing guidance related to stress management and pregnancy. Additionally, 4 studies included a reminder function, which prompted participants to re-engage with the digital intervention if they had not participated for a certain period.

Intervention Delivery Mode

The digital interventions, which were the focus of the studies, were delivered via mobile app in 15 (79%) studies, via the web in 2 (10.5%) studies, and via telemedicine in 2 studies, using video conferencing and WhatsApp Messenger.

Measure of Stress

Stress was measured using various tools. Ten (53%) studies used the Perceived Stress Scale, making it the most frequently used measure. Four (21%) studies used DASS, with one of these also measuring salivary cortisol. Other measures used in the studies included the COVID-19 stress score, prenatal distress questionnaire, Pregnancy Stress Rating Scale-36, Visual Analog Stress Scale, and the Stage of Change in Stress Management, each being used in one study.

Intervention Outcomes

Among the studies, significant changes in intervention outcomes were observed in 10 cases, while in 8 studies, stress levels decreased but were not statistically significant (Table 3). In the study that used a game for stress reduction [28], the DASS results did not exhibit a significant difference, whereas the salivary cortisol results did.

Table 3. Comparisons and outcomes in the included studies.

Author	Intervention group		Control group		Outcomes
	Provided content	Number of participants	Provided content	Number of participants	
Randomized controlled trials					

Balderas-Díaz et al [23]	VivEmbarazo app	15 couples	Routine perinatal care (not mHealth[a] system)	24 couples	• Significant difference in:
					  • Maternal stress (PSS[b]) ↓
					  • Baby weight at birth ↑
					  • Gestational age at birth ↑
					  • Preterm ↓
					• Nonsignificant difference
					  • Maternal depression (EPDS[c])

Lee et al [24]	Self-care for Pregnant Women at Work (SPWW) app	60	Application that only had surveys	56	• At week 4[d], significant difference in:
					  • Pregnancy stress (PDQ[e]) ↓
					  • Pregnancy hassles ↓
					  • Health practice in pregnancy ↑
					• Nonsignificant difference in:
					  • Work stress
					  • Fear of childbirth

Sun et al [25]	Spirits Healing app	74	WeChat health consultations to control attention	84	• At week 8[d], significant difference in:
					  • Depression (EPDS) ↓
					  • Anxiety symptoms (GAD-7[f]) ↓
					  • Position effect ↑
					• Nonsignificant difference in:
					  • Perceived stress (PSS)
					  • Negation effect
					  • Sleep-related problems
					  • Fatigue
					  • Prospective memory
					  • Fear of childbirth

Smith et al [26]	Calm app	33	Routine perinatal care	27	• At day 30[d], significant difference in:
					  • Stress (PSS) ↓
					• Nonsignificant difference in:
					  • Depression (HADS[g])
					  • Anxiety (HADS)
					  • Sleep disturbance

Krusche et al [27]	Website: Be Mindful Online	22 out of 107 respondents who completed the initial survey	Routine perinatal care	50 out of 78 respondents who completed the initial survey	• At day 45[d], significant difference in:
					  • Stress (PSS) ↓
					  • Depression (EPDS) ↓
					  • Pregnancy distress ↓
					• Nonsignificant difference in:
					  • Anxiety (GAD-7)

Dennis-Tiwary et al [28]	ABMT app	15	App with placebo mode	14	• Significant difference in:
					  • Stress (salivary cortisol) ↓
					• Nonsignificant difference in:
					  • Stress (DASS[h])
					  • Anxiety (DASS and HAM-A[i])
					  • Depression (DASS)

Author	Intervention group		Control group		Outcomes
	Provided content	Number of participants	Provided content	Number of participants	
Kia [29]	Mobile-based health educational intervention	40	A PDF file of the educational content	40	• Significant difference in: COVID-19 stress score (CSS-18[j]) ↓

Mauriello et al [30]	iPad- Healthy Pregnancy: Step by Step	169	Brochures named March of Dimes on the target behaviors	166	• At the third trimesterd, nonsignificant difference in:
					  • Stress management
					  • Fruit and vegetable consumption

Chua et al [31]	Parentbot- a digital health care assistant	59 couples	Routine perinatal care	59 couples	• At postpartum months 1 and 3[d], significant difference in
					  • Anxiety (State–Trait Anxiety) ↓
					• Nonsignificant difference in:
					• Stress (PSS)
					  • Depression (EPDS)
					  • Support
					  • Parent-child bonding

Park et al [32]	Mindfulness-based mobile intervention	66	Routine perinatal care, wait-list	67	• At 4 weeks[d], significant difference in:
					  • Anxiety (DASS-21) ↓
					  • Emotional well-being ↑
					  • Maternal-fetal attachment ↑
					• Nonsignificant difference in:
					  • Stress (DASS-21)
					  • Depression (DASS-21)
					  • Postnatal depression (EPDS)

Tandon et al [33]	Personalized stress management	49	Routine perinatal care	51	• At week 1 and postpartum months 1 and 3[d], significant difference in:
					  • Depression (PROMIS[k]) ↓
					  • Perceived stress (PSS-10) ↓
					  • Behavioral activation ↑
					  • Decentering ↑
					  • Mood regulations
					• Nonsignificant difference in:
					  • Anxiety (STAI[l])
					  • Social support

Tian et al [34]	WeChat mini-program (mobile digital platform)	80 couples	Routine perinatal care	80 couples	• At 2 weeks and postpartum week 6[d], significant difference in:
					  • Maternal perceived stress (PSS-10) ↓
					  • Maternal depression (EPDS) ↓
					  • Paternal depression (EPDS) ↓
					  • Mindfulness (FFMQ[m]) ↑
					  • Infant Neuropsychological Development ↑
					• Nonsignificant difference in:
					  • Anxiety (GAD-7)
					  • Symptoms of sleep problems
					  • Fatigue

Author	Intervention group		Control group		Outcomes
	Provided content	Number of participants	Provided content	Number of participants	
Porter et al [35]	Expectful mindfulness app	12 out of 21 participants who completed at least one meditation	Routine perinatal care	247	• At 28 weeks[d], significant difference in stress (PSS) ↓

Jallo et al [36]	Stress coping intervention app	5 out of 15 participants who completed baseline measure	—[n]	—	• Significant change in:
					  • Maternal stress (VASS[o]) ↓
					• Nonsignificant change in:
					  • Perceived stress (PSS)
					  • Stress coping (CSES[p])

Hashemzahi et al [37]	Telemedicine (WhatsApp Messenger)	50	Routine perinatal care	50	• Significant difference in:
					  • Perceived stress (PSS) ↓
					  • Corona disease anxiety (CDAS[q]) ↓

Tsai et al [38]	Web-based antenatal care system	68	Routine perinatal care: face-to-face individual consulting	67	• Significant difference in:
					  • Pregnancy-related stress (PSRS-36[r]) ↓
					  • Self-efficacy (GSE[s]) ↑

Buultjens et al [39]	Perinatal care, education and support (PECS) intervention	40	Routine perinatal care	21	• At 36-38 weeks[d], significant difference in:
					  • Depression (DASS and EPDS)
					• Nonsignificant difference in
					  • Stress and anxiety (DASS)

Kubo et al [40]	Headspace mindfulness app	20	—	—	• Significant change in:
					  • Perceived stress (PSS-10) ↓
					  • Depression (PHQ-8[t]) ↓
					  • Sleep disturbance (PSQI[u]) ↓
					  • Mindfulness (FFMQ) ↑

Barber and Masters-Awatere [41]	Positively Pregnant (PP) mobile app	42	—	—	• Significant change in:
					  • Stress (DASS) ↓
					• Nonsignificant change in:
					  • Depression and anxiety (DASS)
					  • Depression (APDS)

[a]mHealth: mobile health

[b]PSS: Perceived Stress Scale.

[c]EPDS: Edinburgh Postnatal Depression Scale.

[d]In the case of repeated measures, the closest measured values at the end of the intervention were compared.

[e]PDQ: prenatal distress questionnaire.

[f]GAD-7: Generalized Anxiety Disorder.

[g]HADS: Hospital Anxiety and Depression Scale.

[h]DASS: Depression Anxiety Stress Scale.

[i]HAM-A: Hamilton Anxiety Scale.

[j]CSS-18: COVID-19 stress score.

[k]PROMIS: Patient-Reported Outcomes Measurement Information System Depression Scale.

[l]STAI: State-Trait Anxiety Inventory Scale.

[m]FFMQ: Five Facet Mindfulness Questionnaire.

[n]Not applicable.
[o]VASS: Visual Analog Stress Scale.
[p]CSES: Coping Self-Efficacy Scale.
[q]CDAS: Corona Disease Anxiety Scale.
[r]PSRS-36: Pregnancy Stress Rating Scale-36.
[s]GSE: General Self-Efficacy Scale.
[t]PHQ: patient health questionnaire.
[u]PSQI: Pittsburgh Sleep Quality Index.

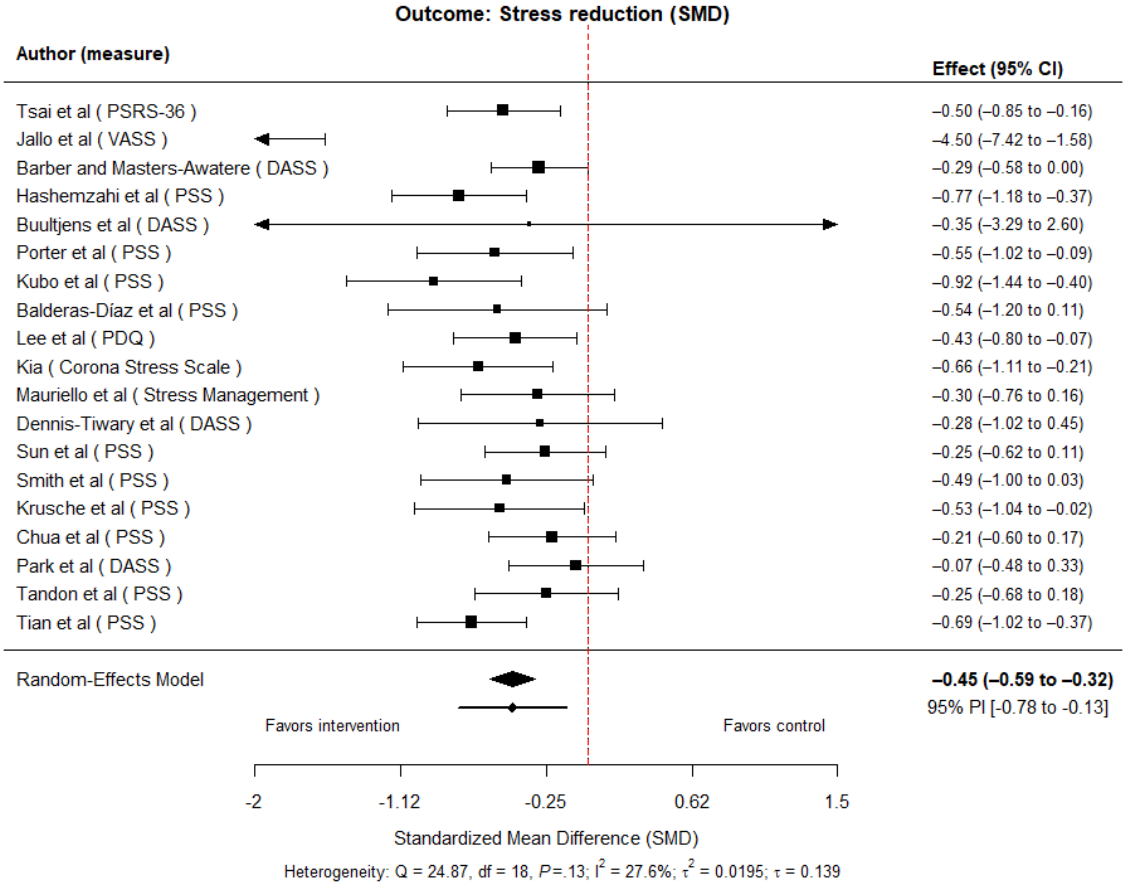
Results of Syntheses

The meta-analysis results of these studies were as follows.

Overall

Across 19 studies, effect sizes were calculated and synthesized. A random-effects meta-analysis using the Hartung–Knapp–Sidik–Jonkman method yielded a pooled average effect of SMD –0.45 (95% CI –0.59 to –0.32), indicating a statistically significant reduction in stress (t_{18}= –6.97, P<.001). Between-study heterogeneity was low to moderate (τ^2=0.0195, τ=0.1395, I^2=27.62%, Q (18)=24.87, P=.13). The 95% PI ranged from –0.78 to –0.13, indicating that the true effects in new settings are expected to remain beneficial overall, although the magnitude of benefit may vary across implementations (Figure 3 [23-41]).

Figure 3. Forest plot of the overall effect of digital interventions on stress reduction [23-41]. DASS: Depression Anxiety Stress Scale; PDQ: prenatal distress questionnaire; PSRS-36: Pregnancy Stress Rating Scale-36; PSS: Perceived Stress Scale; VASS: Visual Analog Stress Scale.



By Strategies

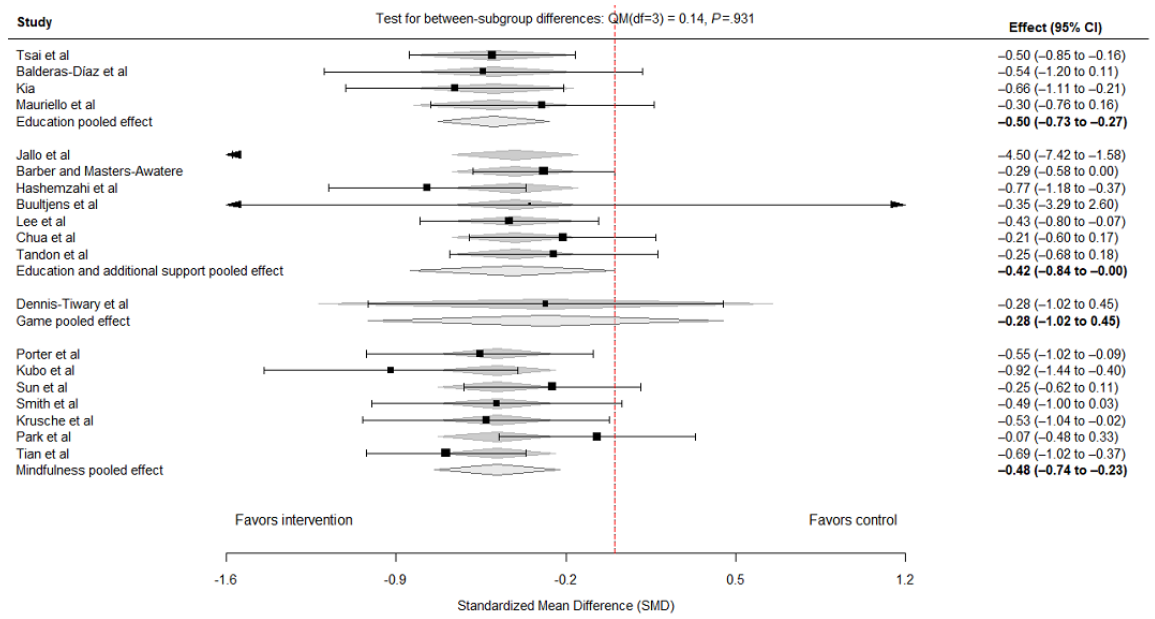
When the intervention effects were categorized by strategies, the pooled estimates were directionally beneficial across subgroups, but statistical certainty varied:

- Education only (k=4): SMD –0.50, 95% CI (–0.73 to –0.27), I^2=0%, τ^2=0
- Education with additional support (k=7): SMD –0.42, 95% CI (–0.84 to 0.00),I^2=52.8%, τ^2=0.059
- Game-based interventions (k=1): SMD –0.28, 95% CI (–1.02 to 0.45),I^2=0%, τ^2=0

- Mindfulness (k=7): SMD –0.48, 95% CI (–0.74 to –0.23),I^2=39.4%, τ^2=0.031

These findings suggest that while several strategies (education-only and mindfulness) show statistically significant average effects, estimates for strategies supported by fewer studies (game-based) or with greater heterogeneity (education with additional support) are less precise and should be interpreted cautiously (Figure 4 [23-41]). A test for between-subgroup differences based on meta-regression showed no statistically significant differences in effect sizes across intervention strategies (Q statistic for moderators: QM_3=0.14; P=.93).

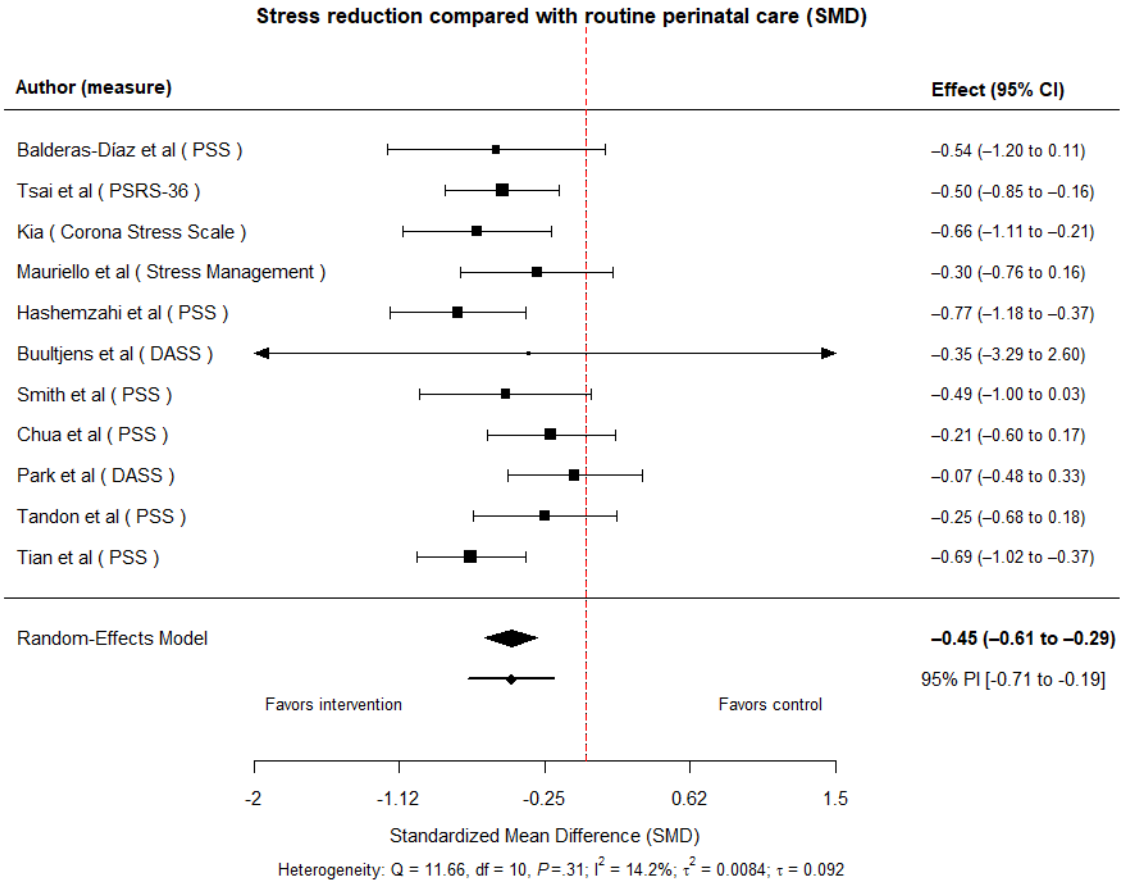
Figure 4. Forest plot for the effects of the intervention by strategies [23-41].



Comparison With Routine Antenatal Care

Studies comparing DHIs with routine antenatal care/usual care showed a significant reduction in stress (pooled average effect SMD −0.45, 95% CI −0.61 to −0.29, t_{10}=−6.27, P<.001). Heterogeneity was low (Q (10)=11.66, P=.31; I^2=14.24%, τ^2=0.0084; Figure 5 [23,26,29-34,37-39]).

Figure 5. Forest plot for the effects compared with routine perinatal care [23,26,29-34,37-39].



Certainty of Evidence

Twelve of the 19 included effect sizes were derived from RCTs. Using the GRADE, the certainty of evidence for stress reduction was rated as moderate, downgraded by one level for risk of bias (1/12 RCTs rated overall high risk, and 2/12 rated some concerns in RoB 2), with no serious concerns identified for inconsistency, indirectness, imprecision, or publication bias.

The certainty of evidence was rated as moderate, downgraded by one level for risk of bias. The pooled effect estimate for RCTs was SMD −0.39 (95% CI −0.52 to −0.26), reflecting a moderate level of certainty in the stress-reducing effect of digital interventions (Multimedia Appendix 3).

Discussion

Principal Findings

Guided by our objectives to characterize antenatal DHIs for stress management, evaluate their effectiveness, and summarize intervention strategies associated with engagement, this systematic review and meta-analysis found that digitally delivered stress-management interventions are generally effective in reducing stress among pregnant women compared with routine care or control conditions. These findings support the growing role of DHIs as a feasible and scalable approach for addressing pregnancy-related stress. Consistent with an umbrella review of digital interventions for perinatal psychological outcomes, digital approaches appear capable of reducing stress symptoms overall, although prior syntheses often pooled stress with other mental health outcomes [14]. They are partially aligned with a recent systematic review and meta-analysis of digitally delivered mindfulness interventions in pregnant women, which reported clear benefits for depression and anxiety and more variable effects for stress [11]. Our findings extend this evidence by focusing specifically on stress as a primary outcome during pregnancy.

When interventions were classified by dominant strategy, mindfulness-based approaches tended to show the most consistent signals of stress reduction. This pattern aligns with earlier reviews of perinatal mindfulness interventions and with emerging trials of digital mindfulness programs tailored to pregnancy [13,42]. Recent randomized trials suggest that mobile-delivered mindfulness programs can reduce depression, underscoring the potential value of skills-based, self-directed content when engagement is maintained [43]. Education-focused programs were also beneficial, particularly when paired with supportive elements such as coaching, peer interaction, or structured follow-up, suggesting that informational content alone may be insufficient for sustained stress reduction. In contrast, gamified interventions were evaluated in only one

study and showed mixed results, with improvements detected in a biological stress indicator but not in self-reported stress, underscoring the need for further research using multiple outcome measures. Importantly, although subgroup-specific point estimates varied across strategies, the formal test for between-subgroup differences was not statistically significant. This indicates insufficient meta-analytic evidence to conclude that any single strategy is superior for stress reduction during pregnancy. Given the limited number of studies within some subgroups and the resulting low statistical power to detect moderator effects, these subgroup patterns should be interpreted cautiously and viewed as exploratory rather than confirmatory.

Beyond intervention strategy, how these programs were delivered also appeared to shape feasibility and engagement. Across studies, most interventions were delivered via mobile apps, with fewer web-based platforms and telemedicine approaches, reflecting contemporary trends in digital health delivery for pregnant populations. The predominance of mobile app–based interventions observed in this review mirrors widespread smartphone use among pregnant women and supports the feasibility of delivering brief, scalable stress-management content within routine antenatal care [16,44]. The increasing dominance of mobile-delivered interventions in more recent studies highlights a shift toward app-based platforms that can leverage a wider range of interactive functions—such as continuous monitoring, automated feedback, and reminders—which may enhance engagement and responsiveness compared with more static web-based programs. While earlier meta-analyses of internet-delivered psychological interventions primarily emphasized benefits for depression and anxiety outcomes [45], our findings suggest that web-based interventions can also contribute to stress reduction during pregnancy, albeit based on a smaller number of trials.

Beyond delivery mode alone, consistent patterns emerged in how interventions were constructed. Most DHIs combined mobile or web-based delivery with self-guided use, supported by functional features such as self-monitoring, automated feedback, reminders, or asynchronous communication, and delivered therapeutic content including psychoeducation, mindfulness practices, relaxation techniques, or cognitive behavioral therapy–based skills. Multi-component designs were the norm rather than the exception, indicating that digital stress-management interventions during pregnancy rarely rely on a single active ingredient. Compared with earlier reviews that grouped interventions primarily by delivery methods or broad perinatal mental health outcomes [13,46], the present review adopts a strategy-oriented perspective that provides a more clinically meaningful framework for understanding how digital interventions may be designed to reduce stress during pregnancy. These findings highlight the need for future research that directly compares single-strategy interventions (such as mindfulness alone or psychoeducation alone) with multi-component interventions, in order to clarify whether additive or synergistic effects contribute to stress reduction in this population.

Interpretation of these findings should consider both between-study variability and study quality. Although between-study variability did not indicate extreme inconsistency, the prediction interval suggests that beneficial effects are likely across future settings, albeit with meaningful variation in magnitude. Reporting prediction intervals alongside CIs helps distinguish the average effect from the distribution of effects expected in real-world implementations [19]. Subgroup findings should therefore be interpreted in the context of overall consistency rather than as definitive evidence of differential effectiveness across strategies [47]. Finally, assessments of small-study effects should be interpreted cautiously, as funnel plot asymmetry may reflect mechanisms beyond publication bias, including true heterogeneity or methodological differences across studies [48].

Clinical Implications

These results suggest that DHIs can serve as a practical adjunct to routine antenatal care by extending access to evidence-based stress-management skills, particularly for women who face barriers to in-person services. Interventions that combine structured mindfulness practice or targeted education with functional support (eg, reminders, progress feedback, brief coaching, or moderated peer support) may be especially useful for maintaining adherence and reinforcing skills [16]. Implementation may be facilitated by integrating these tools into prenatal education pathways and ensuring appropriate guidance for women with elevated distress who may need stepped-up care. Recent trials suggest that mobile-delivered mindfulness may improve depressive symptoms during pregnancy, but sustained engagement and equitable access remain key implementation challenges [43]. From a design perspective, the most promising programs provided clear goals, brief and repeatable programs, and opportunities for personalization. Future DHIs could also explore safe integration of passive sensing or wearable-enabled feedback to tailor content in real time, while ensuring data privacy and minimizing user burden [47,48]. Recent trials also highlight the practical importance of reporting engagement metrics (eg, module completion, frequency of practice, and prompt response) so that implementation decisions can be based not only on efficacy but also on real-world use patterns [43].

Limitations

Several limitations should be noted. The number of trials within some strategy subgroups was small, limiting precision and the ability to detect differential effects. Our restriction to English-language publications and the use of only 4 databases may have missed relevant studies, and most included trials were conducted in high-income countries, which may limit generalizability to low- and middle-income contexts. Outcomes were primarily self-reported, raising the possibility of reporting bias. In addition, incomplete reporting of intervention details and adherence in some trials constrained the interpretation of which components drove benefit. Because the review was not prospectively registered, transparency may be lower than in preregistered reviews; however, we sought to mitigate this limitation by adhering to PRISMA guidance and applying the GRADE approach to characterize certainty of evidence. Moreover, given the rapid evolution of digital perinatal interventions, studies published after our search window may

influence future pooled estimates; ongoing evidence surveillance or living-review approaches may therefore be warranted [14,43].

Conclusions

Overall, this systematic review and meta-analysis provide novel evidence that digitally delivered stress-management interventions can meaningfully reduce stress during pregnancy, highlighting the potential of digital health approaches as a complementary component of antenatal care. Unlike previous reviews that primarily categorized interventions by delivery platform or aggregated stress with other psychological outcomes, this study offers an innovative strategy-based synthesis that disentangles how specific intervention contents and supportive features contribute to stress reduction in pregnant women. By distinguishing intervention strategies rather than technologies alone, the findings advance understanding of what works within digital stress-management programs during pregnancy.

Across studies, the most consistent benefits were observed for mindfulness-based approaches and for educational interventions combined with supportive features such as coaching or peer interaction, underscoring the importance of active skill-building and engagement rather than information provision alone. Importantly, while digital tools are not intended to replace clinician-delivered care, the results indicate that they can extend access to evidence-based coping strategies and provide scalable, low-intensity support that fits within real-world antenatal workflows. As mobile app–based interventions continue to expand, their ability to integrate multiple functions—such as self-monitoring, feedback, and reminders—positions them as particularly feasible and adaptable tools for routine maternity care.

By clarifying the effectiveness of digital stress-management interventions specifically during pregnancy and by framing interventions according to their strategic components, this review contributes actionable insights for clinicians, researchers, and health system planners. The findings support the integration of digital stress-management programs as an adjunct to standard antenatal services, with the potential to improve reach, equity, and continuity of psychosocial support for pregnant women in diverse care settings.

Acknowledgments

We are grateful to Sangkeun Hyon, Dong-Hwan Shin, and Saeam Kim, medical librarians at the Soonchunhyang University Seoul Hospital, for their contributions to database searches and the collection of full text articles for this study. Generative artificial intelligence was used to support English translation and language editing. The authors carefully reviewed and revised all artificial intelligence–assisted text and remain fully responsible for the content of the manuscript.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (number 2023R1A2C1003355 and RS-2023-00239536) as well as by Soonchunhyang University.

Authors' Contributions

JIK contributed to the conceptualization, data curation, and formal analysis, and was involved in writing the original draft, reviewing and editing the manuscript, and providing supervision. JYL contributed to conceptualization and methodology, conducted data curation and formal analysis, and participated in writing the original draft and reviewing and editing the manuscript. SHP contributed to data curation and formal analysis and participated in writing the original draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 Checklist and PRISMA-S Checklist.
[DOCX File , 40 KB - mhealth_v14i1e66267_app1.docx]

Multimedia Appendix 2
Detailed Search Strategies across Databases.
[DOCX File , 16 KB - mhealth_v14i1e66267_app2.docx]

Multimedia Appendix 3
Effect Estimates from Randomized Controlled Trials with GRADE Quality Assessment.
[DOCX File , 25 KB - mhealth_v14i1e66267_app3.docx]

References

1. Maharramova A, Gur EY. The pregnant women's perception of risks and pregnancy stress levels: a cross-sectional study from Turkey. Rev Assoc Med Bras (1992) 2024;70(6):e20231270 [FREE Full text] [doi: 10.1590/1806-9282.20231270] [Medline: 39045948]
2. Williamson SP, Moffitt RL, Broadbent J, Neumann DL, Hamblin PS. Coping, wellbeing, and psychopathology during high-risk pregnancy: a systematic review. Midwifery 2023;116:103556. [doi: 10.1016/j.midw.2022.103556] [Medline: 36427386]
3. Caparros-Gonzalez RA, Torre-Luque ADL, Romero-Gonzalez B, Quesada-Soto JM, Alderdice F, Peralta-Ramírez MI. Stress during pregnancy and the development of diseases in the offspring: a systematic-review and meta-analysis. Midwifery 2021;97:102939. [doi: 10.1016/j.midw.2021.102939] [Medline: 33647755]
4. Cho H, Kim J. Moderating effect of general self-efficacy on the relationship between pregnancy stress, daily hassles stress, and preterm birth risk in women experiencing preterm labor: a cross-sectional study. J Korean Acad Nurs 2024;54(3):329-339. [doi: 10.4040/jkan.24008] [Medline: 39248420]
5. Missler M, Donker T, Beijers R, Ciharova M, Moyse C, de Vries R, et al. Universal prevention of distress aimed at pregnant women: a systematic review and meta-analysis of psychological interventions. BMC Pregnancy Childbirth 2021;21(1):276 [FREE Full text] [doi: 10.1186/s12884-021-03752-2] [Medline: 33794828]
6. Nasrollahi M, Ghazanfar Pour M, Ahmadi A, Mirzaee M, Alidousti K. Effectiveness of mindfulness-based stress reduction on depression, anxiety, and stress of women with the early loss of pregnancy in southeast Iran: a randomized control trial. Reprod Health 2022 Dec 29;19(1):233 [FREE Full text] [doi: 10.1186/s12978-022-01543-2] [Medline: 36581926]
7. Corrigan L, Moran P, McGrath N, Eustace-Cook J, Daly D. The characteristics and effectiveness of pregnancy yoga interventions: a systematic review and meta-analysis. BMC Pregnancy Childbirth 2022;22(1):250 [FREE Full text] [doi: 10.1186/s12884-022-04474-9] [Medline: 35337282]
8. Huo L. Effects of prenatal psychotherapies and psychosocial interventions on depressive symptoms, anxious symptoms and stress: a systematic review and network meta-analysis. Front Psychiatry 2025;16:1624924 (forthcoming) [FREE Full text]
9. Sîrbu V, David OA. Efficacy of app-based mobile health interventions for stress management: a systematic review and meta-analysis of self-reported, physiological, and neuroendocrine stress-related outcomes. Clin Psychol Rev 2024;114:102515 [FREE Full text] [doi: 10.1016/j.cpr.2024.102515] [Medline: 39522422]
10. Linardon J, Firth J, Torous J, Messer M, Fuller-Tyszkiewicz M. Efficacy of mental health smartphone apps on stress levels: a meta-analysis of randomised controlled trials. Health Psychol Rev 2024;18(4):839-852 [FREE Full text] [doi: 10.1080/17437199.2024.2379784] [Medline: 39041586]
11. Mefrouche ML, Siegmann E, Böhme S, Berking M, Kornhuber J. The effect of digital mindfulness interventions on depressive, anxiety, and stress symptoms in pregnant women: a systematic review and meta-analysis. Eur J Investig Health Psychol Educ 2023;13(9):1694-1706 [FREE Full text] [doi: 10.3390/ejihpe13090122] [Medline: 37754461]
12. Han Y, Tian Q, Xu M, Zhao W, Wang Z, Zhang W. Effects of nurse-led e-health interventions on the health-related outcomes of pregnant women: a systematic review. J Clin Nurs 2025;34(1):88-107. [doi: 10.1111/jocn.17560] [Medline: 39568149]
13. Stentzel U, Grabe HJ, Schmidt S, Tomczyk S, van den Berg N, Beyer A. Mental health-related telemedicine interventions for pregnant women and new mothers: a systematic literature review. BMC Psychiatry 2023;23(1):292 [FREE Full text] [doi: 10.1186/s12888-023-04790-0] [Medline: 37118689]
14. Lau Y, Chew HSJ, Ang WHD, Ang WW, Yeo CY, Lim GZQ, et al. Effects of digital health interventions on the psychological outcomes of perinatal women: umbrella review of systematic reviews and meta-analyses. Health Psychol Rev 2024;18(2):229-254 [FREE Full text] [doi: 10.1080/17437199.2023.2185654] [Medline: 36919443]
15. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
16. Paganini S, Meier E, Terhorst Y, Wurst R, Hohberg V, Schultchen D, et al. Stress management apps: systematic search and multidimensional assessment of quality and characteristics. JMIR Mhealth Uhealth 2023;11:e42415 [FREE Full text] [doi: 10.2196/42415] [Medline: 37642999]
17. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
18. Rethlefsen ML, Kirtley S, Waffenschmidt S, Patricia Ayala A, Moher D. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021;10(1). [doi: 10.31219/osf.io/sfc38]
19. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
20. 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. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]

21. Borenstein M. How to understand and report heterogeneity in a meta-analysis: the difference between I-squared and prediction intervals. Integr Med Res 2023;12(4):101014 [FREE Full text] [doi: 10.1016/j.imr.2023.101014] [Medline: 38938910]
22. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard Dersimonian-Laird method. BMC Med Res Methodol 2014;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
23. Balderas-Díaz S, Rodríguez-Fórtiz MJ, Garrido JL, Bellido-González M, Guerrero-Contreras G. A psycho-educational intervention programme for parents with SGA foetuses supported by an adaptive mHealth system: design, proof of concept and usability assessment. BMC Med Inform Decis Mak 2022;22(Suppl 4):291 [FREE Full text] [doi: 10.1186/s12911-022-02036-9] [Medline: 36357878]
24. Lee Y, Kim S, Choi S. Effectiveness of mobile-based intervention self-care for pregnant women at work: a randomized controlled trial. J Occup Health 2023;65(1):e12402 [FREE Full text] [doi: 10.1002/1348-9585.12402] [Medline: 37144249]
25. Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F. Effectiveness of smartphone-based mindfulness training on maternal perinatal depression: randomized controlled trial. J Med Internet Res 2021;23(1):e23410 [FREE Full text] [doi: 10.2196/23410] [Medline: 33502326]
26. Smith R, Mahnert ND, Foote J, Saunders KT, Mourad J, Huberty J. Mindfulness effects in obstetric and gynecology patients during the coronavirus disease 2019 (COVID-19) pandemic: a randomized controlled trial. Obstet Gynecol 2021;137(6):1032-1040 [FREE Full text] [doi: 10.1097/AOG.0000000000004316] [Medline: 33957663]
27. Krusche A, Dymond M, Murphy SE, Crane C. Mindfulness for pregnancy: a randomised controlled study of online mindfulness during pregnancy. Midwifery 2018;65:51-57. [doi: 10.1016/j.midw.2018.07.005] [Medline: 30099285]
28. Dennis-Tiwary TA, Denefrio S, Gelber S. Salutary effects of an attention bias modification mobile application on biobehavioral measures of stress and anxiety during pregnancy. Biol Psychol 2017;127:148-156 [FREE Full text] [doi: 10.1016/j.biopsycho.2017.05.003] [Medline: 28478138]
29. Kia ZR. The effect of educational intervention with mobile health technology on COVID-19 induced stress among pregnant women: a randomized controlled trial. J Midwifery Reprod Health 2023;11(3):3782-3793 [FREE Full text] [doi: 10.22038/JMRH.2022.66480.1940]
30. Mauriello LM, Van Marter DF, Umanzor CD, Castle PH, de Aguiar EL. Using mHealth to deliver behavior change interventions within prenatal care at community health centers. Am J Health Promot 2016;30(7):554-562. [doi: 10.4278/ajhp.140530-QUAN-248] [Medline: 26305603]
31. Chua JYX, Choolani M, Chee CYI, Yi H, Chan YH, Lalor JG, et al. The effectiveness of parentbot - a digital healthcare assistant - on parenting outcomes: a randomized controlled trial. Int J Nurs Stud 2024;160:104906. [doi: 10.1016/j.ijnurstu.2024.104906] [Medline: 39305680]
32. Park S, Cho HY, Park JY, Chung K, Jhung K. Development and evaluation of a mindfulness-based mobile intervention for perinatal mental health: randomized controlled trial. J Med Internet Res 2025;27:e56601 [FREE Full text] [doi: 10.2196/56601] [Medline: 39823585]
33. Tandon SD, Moskowitz JT, Edwards RC, Zhang Y, Giase G, Sinche B, et al. Effects of a personalized stress management intervention on maternal mental health: a randomized clinical trial. Arch Womens Ment Health 2025;28(6):1585-1595. [doi: 10.1007/s00737-025-01619-5] [Medline: 40960524]
34. Tian Y, Ma R, Cui N, Wang J, Huang Y, Guo K, et al. Effects of digital mindfulness training for couples on psychological distress and infant neuropsychological development: randomized controlled trial. J Med Internet Res 2025;27:e77260 [FREE Full text] [doi: 10.2196/77260] [Medline: 41271207]
35. Porter AC, Hunter S, Noonan K, Hoffman MC. A mindfulness application for reducing prenatal stress. J Midwifery Womens Health 2022;67(4):442-447 [FREE Full text] [doi: 10.1111/jmwh.13359] [Medline: 35403807]
36. Jallo N, Thacker LR, Menzies V, Stojanovic P, Svikis DS. A stress coping app for hospitalized pregnant women at risk for preterm birth. MCN Am J Matern Child Nurs 2017;42(5):257-262. [doi: 10.1097/NMC.0000000000000355] [Medline: 28817447]
37. Hashemzahi M, Khayat S, Khazaeian S. Effect of COVID-19 self-care training via telemedicine on perceived stress and corona disease anxiety in pregnant women: a quasi-experimental study. J Midwifery Reprod Health 2022;10(1):3066-3074 [FREE Full text] [doi: 10.22038/jmrh.2021.60589.1728]
38. Tsai Y, Hsu Y, Hou T, Chang C. Effects of a web-based antenatal care system on maternal stress and self-efficacy during pregnancy: a study in Taiwan. J Midwifery Womens Health 2018;63(2):205-213. [doi: 10.1111/jmwh.12685] [Medline: 29533525]
39. Buultjens M, Gill J, Fielding J, Lambert KA, Vondeling K, Mastwyk SE, et al. Maternity care during a pandemic: can a hybrid telehealth model comprising group interdisciplinary education support maternal psychological health? Women Birth 2023;36(3):305-313 [FREE Full text] [doi: 10.1016/j.wombi.2022.09.007] [Medline: 36184532]
40. Kubo A, Aghaee S, Kurtovich EM, Nkemere L, Quesenberry CP, McGinnis MK, et al. mHealth mindfulness intervention for women with moderate-to-moderately-severe antenatal depressive symptoms: a pilot study within an integrated health care system. Mindfulness (N Y) 2021;12(6):1387-1397 [FREE Full text] [doi: 10.1007/s12671-021-01606-8] [Medline: 33723491]

41. Barber CC, Masters-Awatere B. Positively pregnant: development and piloting of a mobile app for social and emotional well-being in pregnancy. Appl Psychol Health Well Being 2022;14(4):1255-1272. [doi: 10.1111/aphw.12333] [Medline: 34959260]
42. Lucena L, Frange C, Pinto ACA, Andersen ML, Tufik S, Hachul H. Mindfulness interventions during pregnancy: a narrative review. J Integr Med 2020;18(6):470-477. [doi: 10.1016/j.joim.2020.07.007] [Medline: 32798196]
43. Kim M, Park JY, Park S, Chung K, Cho HY, Do G, et al. Effectiveness of a mindfulness-based mobile intervention for improving perinatal mental health and reducing depression during pregnancy: randomized controlled trial. J Med Internet Res 2025;27:e75630 [FREE Full text] [doi: 10.2196/75630] [Medline: 41037777]
44. Nabovati E, Farzandipour M, Vahedpoor Z, Akbari H, Anvari S, Sharif R, et al. Pregnant women's use and attitude toward mobile phone features for self-management. BMC Med Inform Decis Mak 2023;23(1):77 [FREE Full text] [doi: 10.1186/s12911-023-02172-w] [Medline: 37101302]
45. Neo HS, Tan JH, Ang WHD, Lau Y. Internet-delivered psychological interventions for reducing depressive, anxiety symptoms and fear of childbirth in pregnant women: a meta-analysis and meta-regression. J Psychosom Res 2022;157:110790. [doi: 10.1016/j.jpsychores.2022.110790] [Medline: 35367919]
46. Nair U, Armfield NR, Chatfield MD, Edirippulige S. The effectiveness of telemedicine interventions to address maternal depression: a systematic review and meta-analysis. J Telemed Telecare 2018;24(10):639-650. [doi: 10.1177/1357633X18794332] [Medline: 30343660]
47. Borenstein M, Higgins JPT. Meta-analysis and subgroups. Prev Sci 2013;14(2):134-143. [doi: 10.1007/s11121-013-0377-7] [Medline: 23479191]
48. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;343:d4002. [doi: 10.1136/bmj.d4002] [Medline: 21784880]

Abbreviations

DASS: Depression Anxiety Stress Scale
DHI: digital health intervention
GRADE: Grading of Recommendations, Assessment, Development and Evaluation
MeSH: Medical Subject Headings
PICOS: population, intervention, comparison, outcome, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension
RoB 2: revised Cochrane risk-of-bias tool
ROBINS-I: risk of bias in nonrandomized studies of interventions
SMD: standardized mean difference

Edited by S Brini; submitted 09.Sep.2024; peer-reviewed by E Motrico, A Finlay; comments to author 12.Mar.2025; revised version received 28.Dec.2025; accepted 28.Dec.2025; published 26.Jan.2026.

Please cite as:
Kim JI, Lee JY, Park SH
Evaluating the Effectiveness of Digital Interventions for Stress Management in Pregnant Women: Systematic Review and Meta-Analysis
JMIR Mhealth Uhealth 2026;14:e66267
URL: https://mhealth.jmir.org/2026/1/e66267
doi:10.2196/66267
PMID:

Features of Mobile Health Apps for Tobacco Cessation That Appeal to Black Adults Who Use Tobacco Products: Focus Group Study

Sonia A Clark[1,2], MHA; Remi Philips[2], BS; Christine E Kistler[3], MD, MASc; Adam O Goldstein[1,2], MD, MPH; Chineme Enyioha[1,2], MD, MPH

[1]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Department of Family Medicine, University of North Carolina at Chapel Hill, 590 Manning Dr, Chapel Hill, NC, United States

[3]Division of Geriatrics, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

Corresponding Author:
Chineme Enyioha, MD, MPH
Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Abstract

Background: Mobile health (mHealth) interventions show promise in supporting tobacco cessation. However, Black adults who use tobacco products are not well represented in mHealth studies for tobacco cessation, and their preferred features of mHealth apps are not well known. Identifying types of mHealth app features for tobacco cessation preferred by Black adults is critical to developing a culturally adapted app, with increased uptake by the target population.

Objective: The goal of this study was to explore culturally relevant preferences for features of smoking cessation mHealth apps among Black adults who use tobacco products.

Methods: A comprehensive list of features of mHealth apps for tobacco cessation was developed based on previous research and a review of existing mHealth literature. Through a content analysis, this list was divided into subgroups and used to develop a focus group guide. We recruited participants from Instagram, a social media platform. Eligible focus group participants included people who reported current use of a tobacco product, identified as being African American or Black, were 21 years old or older, and had access to Wi-Fi or the internet. Participants had to indicate interest in the use of an mHealth app for tobacco cessation. Participants discussed their opinions about different app features, including what features they felt would increase the use of an app by Black adults. Recordings from the focus groups were transcribed and coded deductively and inductively. We conducted a thematic content analysis of the resulting transcripts.

Results: Forty adults aged 21 - 69 (mean 43, SD 13.6) years participated in 8 focus groups. Fifty-seven percent were female, and 88% endorsed current cigarette use. Four central themes that represented app features emerged. (1) Participants wanted representation and inclusivity through personalization and featuring people with similar lived experiences, including representative images and relevant health information. (2) Participants desired the app to feature a diversity of experiences such as testimonials from individuals from different backgrounds rather than solely focusing on racial identity or excessive targeting of the Black community. (3) Participants desired accountability through trusted connections with health care professionals and other support groups within the app, as well as app tracking capability. (4) Encouragement and motivation were more salient incentives than monetary rewards.

Conclusions: Black adults who use tobacco products prefer a tobacco cessation app with features that are inclusive, relatable, supportive, and motivating. These findings can serve as the groundwork for the development of an mHealth app that will appeal to Black adults, potentially leading to increased app use, successful cessation, and health equity.

(JMIR Mhealth Uhealth 2026;14:e63340) doi:10.2196/63340

KEYWORDS

mHealth; tobacco cessation; preferred features; mobile health app; representation; inclusivity; African American; Black adult tobacco users; mobile health

Introduction

In the United States, certain racial minorities, including Black adults, face a disproportionate burden of negative health effects from tobacco use [1]. Compared to other racial and ethnic groups, Black adults who use tobacco products are more likely to experience adverse health issues because of their tobacco use, such as coronary heart disease, and have a higher morbidity

and mortality rate from other tobacco-related health conditions [2,3]. For instance, compared with Whites, Black adults have a higher mortality rate from cancer (221 vs 186 per 100,000 population) [4]. Despite starting to smoke at an older age [5] and being more likely to make a quit attempt than any other racial group, Black adults who use tobacco products are less likely to successfully quit than other racial groups [6]. Black adults also have lower odds of both quitting and using Food and Drug Administration cessation medications, despite having greater odds of using behavioral counseling/self-help materials [7].

Novel interventions, such as mobile health (mHealth) interventions, show promise in supporting tobacco cessation for Black adults. mHealth apps have previously been successfully used with Black adults to help manage conditions such as diabetes and obesity [8,9]. Compared to White adults, Black adults are more likely to use apps to inform their medical decisions [10]. mHealth apps can also increase access for Black adults who use tobacco products by reducing barriers to receiving care, like cost and transportation logistics [11,12]. One study examining willingness to participate in mHealth app research for weight management among Black women found high receptivity, frequent smartphone usage, and existing use of smartphones to access information on health and wellness, suggesting mHealth apps could be used to empower and promote participation [13].

Although a promising route to increase tobacco cessation, mHealth apps should be culturally adapted to address unique factors that impact Black adults who use tobacco products. Black adults use menthol products and smoke cigars at higher rates [14], smoke fewer cigarettes per day [15], and may be more likely to use tobacco products as a result of stress related to racial harassment [16]. One study examining differences in addiction and quitting experiences concluded that these differences may suggest that Black adults who use tobacco products experience a different “quitting process” [15] as compared to White smokers. However, Black adults who use tobacco have not been well represented in prior mHealth studies for tobacco cessation, and their preferred features of mHealth apps are not well known. In addition, research has shown that tailoring interventions in ways that incorporate values and cultural experiences for patients, especially patients from minority backgrounds, can increase engagement with the intervention [17]. For example, visual representation of an individual’s culture in an intervention, delivery or promotion of an intervention by an individual from the target population, or the use of language that the target audience can relate to can enhance the appeal of that intervention [17,18]. A critical step to developing a tobacco cessation mHealth app that Black adults are receptive to and are likely to use is to identify desirable app features.

Furthermore, the extent to which preferred features satisfy key components of the self-determination theory is also important. According to the self-determination theory, autonomy, competence, and relatedness are key factors necessary for intrinsic motivation as well as maintenance of behavior change [19]. Autonomy, described as the intrinsic need to regulate one’s own behavior without the influence of external factors, has been found to be effective in tobacco cessation [20]. Competence, also described as the need for an individual to feel confident and capable as it relates to behavior change, can be supported by app features such as an app’s ability to provide information about tobacco cessation or provision of information about a user’s progress or behavior [21]. Relatedness is described as connecting with others and can be achieved by an app that allows for social support or interaction with others [21]. Each component of the self-determination theory plays an important role in cessation-related behavior [20,22]. The goal of this study was to identify features of mHealth apps for tobacco cessation that Black adults who use tobacco products would like to see in an app and what features will make the app more inclusive [19-22].

Methods

Study Design

Between June and July of 2023, we conducted 8 focus groups with 40 Black adults who use tobacco products. Participants were recruited using the social media platform Instagram. After completing a brief screener and obtaining consent during a brief screening call, eligible participants were purposively invited to participate in focus groups. Focus groups were stratified by age groups (21 - 34, 35 - 54, and ≥55 years old) to optimize representation.

Participants

Eligible participants included Black adults who were 21 years old or older, who reported current use of one or more tobacco products in the past 30 days, spoke English, and were current US residents at the time of recruitment. Exclusion criteria included lack of access to the internet and lack of interest in the use of technology for smoking cessation.

Focus Group Guide Development

To identify which features to discuss during focus groups, we developed a comprehensive list of features of mHealth apps for tobacco cessation using previous research and a review of existing mHealth literature [23-31]. The mHealth apps involved were designed based on behavior change techniques and theories, such as the principles of acceptance and commitment therapy [24,26], persuasive technology for behavior change [27], or followed the US clinical practice guidelines [28,31]. One of the apps was developed by the National Cancer Institute [28], and another was developed to target young adults who are motivated to quit [27]. We then divided this list into different domains to explore with participants. Initial mHealth app feature domains included content, user experience, personalization, tracking, privacy and security, connections or support, and inclusivity. Within each domain, several features were included to discuss with participants, with the primary focus being whether these features were inclusive and preferred by Black adults who use tobacco products. Using guidance from experts, the resulting domains and features were organized and distilled. The list and resulting domains were further refined through focus groups using the focus group guide.

Focus Group Procedures

Using the list of domains and features of mHealth apps, the research team developed the focus group guide to address the research question: What features in mHealth apps for tobacco cessation are preferred by Black adults? The interview guide included questions and probes about participants' experience with tobacco use and cessation, and experience with mHealth apps. Next, the guide listed each of the domains and features with a corresponding question about what features they preferred and a final question to get feedback about other design features or advice participants had. The semistructured focus groups were conducted using Zoom videoconferencing software to allow for participation from different locations within the United States, and each focus group lasted an hour [32].

Analysis

The resulting focus group recordings were professionally transcribed, and transcripts were uploaded into Atlas.ti, a qualitative data analysis software (version 23). Using the focus group guide, an initial codebook was developed. The qualitative analysis team comprised 3 researchers (CE, SAC, and RP), who each had prior experience with qualitative coding and received formal qualitative and mixed methods training at the University of North Carolina at Chapel Hill. After familiarizing themselves with the codebook, 2 researchers (SAC and RP) independently coded a randomly selected transcript with the guidance of the senior researcher (CE). To establish the rigor of the analysis, we applied multiple methods to strengthen the credibility, dependability, and confirmability. First, we documented codebook revisions and applied analytic triangulation with multiple coders who reviewed transcripts independently and participated in consensus meetings to review discrepancies, add new codes, and refine definitions as needed. To assess interrater reliability, Krippendorff α was used, and the initial interrater reliability was calculated to be 0.769, indicating substantial agreement. After resolving coding conflicts, a different transcript was then coded, and "near perfect" agreement, 0.9, was reached. Upon reaching this threshold of high agreement, the team recoded the initial transcript and coded the remaining transcripts with the finalized codebook.

Furthermore, the focus groups were conducted with cultural sensitivity. For instance, the main facilitator of the focus groups was a member of the research team and is also a part of the Black community, as well as familiar with participants' cultural values [32]. We did not conduct member check-in to reduce the burden on participants and concerns for confidentiality. However, the data were presented to the larger research team, and findings were discussed in detail [33].

Using thematic content analysis, the research team discussed and identified recurring themes and concepts across the transcripts [34]. The team determined that thematic saturation was achieved after the completion of 8 focus groups [35].

Ethical Considerations

This study was approved by the institutional review board at the University of North Carolina at Chapel Hill (24 - 0798). All study participants were informed about the study aims and significance, as well as the methods. Participants gave their informed consent, retained the rights to withdraw from the study at any time, and were informed of the confidentiality and anonymity of the data and the goal to publish findings from the study. All participants received a US $50 gift card for their participation in the study.

Results

Sociodemographic and Tobacco Use Data

Forty adults aged 21-69 years, with an average age of 43 (SD 13.6) years, participated in 8 focus groups (Table 1). Most participants were female, and the majority were non-Hispanic. Many participants endorsed current cigarette use. Fifty-five percent (n=22) reported use of electronic cigarettes or other vaping devices, while 40% (n=16) endorsed use of cigars.

Table . Demographics and tobacco use characteristics of participants (N=40).

Characteristics	Value
Age (years), mean (SD)	43 (13.6)
Sex (female), n (%)	23 (57)
Race, n (%)	
Black/African American	36 (90)
Black/African American and some other race, ethnicity, or origin	4 (10)
Ethnicity (non-Hispanic), n (%)	37 (92)
Sexual orientation, n (%)	
Straight	33 (82)
Gay or lesbian	2 (5)
Bisexual	5 (13)
Current cigarette use, n (%)	
Yes	35 (88)
Current other tobacco product use, n (%)	
Yes	31 (78)
Electronic cigarettes, e-cigarettes, or other vaping devices	22 (55)
Cigars, for example, little cigars, cigarillos, or large cigars	16 (40)
Smokeless tobacco, for example, chewing tobacco, snuff, or snus	6 (15)
Water-pipe tobacco or hookah	5 (13)
Other types of oral nicotine products, such as Velo lozenges, Rogue tablets, Lucy gum, or Pixotine toothpicks	1 (3)
Other	3 (8)
None of the above	9 (23)

Four central themes emerged about what mHealth app features were desirable by Black adults who use tobacco products. Each theme had 2 - 3 subthemes that represented different domains of mHealth app features. See Table 2.

Table . Themes and subthemes with domains of mobile health app features represented.

Themes	Subthemes	Domain of app feature
Participants wanted representation and inclusivity through personalization and featuring people with similar lived experiences	• Inclusion of images to represent the target population • Personalization of the app • Inclusion of content about relevant health risks	• Representation and inclusivity • User experience • Content and inclusivity
Excessive targeting of the Black community was unappealing, and participants desired the app to feature a diversity of experiences rather than solely focusing on racial identity	• Inclusion of testimonials from people with different backgrounds • Outreach in the Black community	• Content • Social support
Participants desired accountability through trusted connections, and app tracking capability	• Connection with health professionals • Supportive anonymous community within the app • App tracking capability	• Professional support • Social or peer support • Tracking/user experience
Encouragement and motivation were more salient incentives than monetary rewards	• Nonmonetary rewards • Provision for diverse reward options	• Gamification and rewards

Theme 1: Participants Wanted Representation and Inclusivity Through Personalization and Featuring People With Similar Lived Experiences

Content and Inclusivity: Inclusion of Images to Represent the Target Population

When asked how to make the app more inclusive of Black adults, several participants emphasized the need to create an app that features Black adults.

> *If we get on the app and we see a lot of us, then that will make us wanna use it more.* [Focus group 1, Interviewee 1]

> *You gonna get more people who's like, "Hey, that applies to me. First of all, they look like me. They talk like me...They kinda think like me."* [Focus group 3, Interviewee 6]

One person lamented that it was not common to see themselves represented in existing apps.

> *You know, like I need to see people of color, you know..., connecting my eyes because it's not [in] a lot of apps out here...* [Focus group 6, Interviewee 2]

Testimonials from fellow Black individuals were seen as a way to make people want to use the app:

> *...having testimonials of other African-Americans, 'cause that would like make me at least feel more comfortable to even wanna try the app. Just seeing that other people that look like me have used it and have something positive to say.* [Focus group 4, Interviewee 3]

User Experience: Personalization of the App

Some believed that inclusivity could be achieved through personalization of tobacco use and cessation goals.

> *...[to] feel like you [are] talking to somebody that you know...maybe you could pick up from the personalized questions in beginning.* [Focus group 6, Interviewee 2]

Participants also believed personalization could be achieved through in-app customization, which can allow people to have their goals at the forefront.

> *...you should be able to customize [the app], you know, or personalize your experience with certain personal goals and..., based on your smoking habits...* [Focus group 4, Interviewee 2]

> *..Then this goal is set to be like the basis of the notification and the recommendations you receive.* [Focus group 4, Interviewee 2]

Content and Inclusivity: Inclusion of Content About Relevant Health Risks

The inclusion of content about their health risks as Black adults who use tobacco products was also suggested:

> *Include like the health aspect of it, like what smoking contributes to, like, just to add the statistics like, you know, African Americans,...make sure that you mention more of how it affects us as-as a culture.* [Focus group 7, Interviewee 5]

With personalized content, participants hope to find stories to relate to.

> *Personalize it because everyone has their own story. So, if you can find the testimonials that can represent you, that would work.* [Focus group 1, Interviewee 2]

Some indicated that having personalized content would be preferable because they have different experiences and different stressors.

> *...involve Black people [36], we have shared experiences that only [we] can understand...Like a Black mother or Black woman—if I need help, she could like advise me and pull me through.* [Focus group 4, Interviewee 4]

> *I think people that use cigarettes as a coping mechanism...I have—for example, stressors in the neighborhood...* [Focus group 3, Interviewee 2]

Theme 2: Excessive Targeting of the Black Community Was Unappealing, and Participants Desired the App to Feature a Diversity of Experiences Rather Than Solely Focusing on Racial Identity

Content: Inclusion of Testimonials From People With Different Backgrounds

While many participants voiced that the app should offer representation and inclusivity where Black adults are represented, many were adamant that the app should not focus on Black adults solely.

> *I don't wanna feel like it's directed right at me 'cause I'm Black. 'Cause I'm more than just Black.* [Focus group 2, Interviewee 2]

Rather, participants voiced that they wanted an inclusive app open to everyone struggling with smoking.

> *It should be point to everyone with smokes, despite religion, despite your-your race, despite whoever you are.* [Focus group 1, Interviewee 3]

> *...provide practical solutions accessible to users from various income levels. You know, regardless of-regardless of whether you Black or not.* [Focus group 4, Interviewee 2]

Participants suggested that one way to bring diversity into the app is to include testimonials from people from different backgrounds.

> *I would like to hear,...different people's testimonials...because we all have different stories, you know, and different reasons for doing the things that we do.* [Focus group 1, Interviewee 1]

Having the app feature real stories was viewed as a way to motivate app use:

> *...if you share like a real story of somebody that, you know, suffered bad consequences for smoking for not*

so long, that will probably motivate me to use the app. [Focus group 7, Interviewee 7]

Social Support: Outreach in the Black Community

Finally, some participants suggested conducting outreach at events within the Black community to promote the app and collaborate with respected figures:

...have it displayed in a way that's attractive to us, and then also present it in the areas that we are...events that we go to. [Focus group 6, Interviewee 4]

You need to get...somebody famous. I'mma just use Diana Ross just for name sake, right? Say she used to smoke. You get her to talk about it. We all know, in the African American community, word of mouth is the most powerful thing we have. [Focus group 3, Interviewee 6]

Some participants suggested social media marketing:

...if I have like a good TikTok video of someone that's like a Black that's...reviewing or giving me something to pull me in, that's definitely gonna make me go and download it. [Focus group 6, Interviewee 4]

Another participant reflected on their experience of trying an app because it was recommended by an influencer:

I actually got to know about it by...[a] Black female YouTuber and that's how I got interested...

[Focus group 4, Interviewee 4]

Theme 3: Participants Desired Accountability Through Trusted Connections and App Tracking Capability

Professional Support: Connection With Health Professionals

When asked whether the app should offer connections to other app users, doctors, or other health care providers, participants expressed their desire to connect in order to promote accountability. Some participants indicated talking to professionals would help keep them accountable to their goals, either by providing advice, "that would seem pretty nice to be able to...have a coach that, you know, that you can go to for advice or assistance..." [Focus group 1, Interviewee 1] or by monitoring their progress, "I think you should have some kind of feature that to share your usage with your physician and stuff like that or your coach" [Focus group 4, Interviewee 2].

Social Support: Supportive Anonymous Community Within the App

When asked about connections with other users of the app, most participants indicated interest in seeking supportive accountability within the app:

I think it will be better to have someone that, you know—'cause my friend doesn't smoke, so someone that understands. But to be able to hold you accountable. [Focus group 1, Interviewee 1]

Participants expressed a desire to decrease loneliness by connecting with others who understand what they are going through:

I find that quitting smoking is like the loneliest thing I've ever dealt with in my life...[If] there's like a little message board type of thing inside the app where people can just write in,...what they're going through. [Focus group 1, Interviewee 2]

Most participants preferred the community within the app to be anonymous and were interested in anonymous interactions through groups, games, and a leaderboard, "to kind of promote data security...the information of the user should be de-identified or...be on the app as anonymous user" [Focus group 4, Interviewee 2] and "...in terms of being able to see what might be called like a leaderboard or a scoreboard...it can kind of make it...a bit more interactive" [Focus group 7, Interviewee 3].

Although most participants wanted a quit buddy within the app, they did not want this quit buddy to be a family member or friend. A few people noted that friends and family can be triggers for smoking and they are not understanding, "because sometimes family and friends can be triggers to smoking" [Focus group 8, Interviewee 3] and "sometimes like she said, you can't really talk to family or friends about it, 'cause they don't—they don't do it 'cause they don't understand it..." [Focus group 1, Interviewee 1]. While most participants did not want their friends and family connected to the app, some expressed that it would be helpful to have them understand and provide accountability: "It'll give them an insight of what we're going through" [Focus group 8, Interviewee 1] and "I think it's a good idea to have, help because it could help you track your health and with how many smoking occurrences have you smoked per smoke a day, and can get your family involved and friends" [Focus group 8, Interviewee 7]. When asked if they wanted the app to connect with social media, some participants expressed fear of judgment and mean comments from other social network apps.

You know, there's some nasty comments, tellin' you to just get over it. [Focus group 5, Interviewee 6]

I don't want that stuff flashed across Facebook and-and everybody to know what I'm tryin' to do or attemptin' to do because there's a lotta unnecessary opinions. [Focus group 5, Interviewee 6]

Tracking and User Experience: App Tracking Capability

When asked about app tracking capabilities, most participants noted that they want to be able to track their usage to keep them accountable.

Say, I might smoke this many a day. How 'bout the next day—and-and-and a tracker that helps me keep accountability. [Focus group 5, Interviewee 6]

A few participants noted that tracking their triggers through the app would allow them to learn what to avoid.

It help you identify your triggers, and you'll see,...what things cause you to smoke more so you could, like, lean away from it or stay from it,

> *dependin' on what it is. So, I think it's [an] excellent idea.* [Focus group 3, Interviewee 6]

Some participants expressed a desire to have frequent notification reminders to support accountability.

> *So I need an app that's going to be a constant reminder, almost an annoyance, you know, and - and to help me.* [Focus group 8, Interviewee 2]

Most participants indicated that accountability could be supported through reminders of their goals.

> *The daily goals are there, the notifications are there to remind you of your goals and all, so, uh, that should be helpful.*
>
> *[Focus group 4, Interviewee 5]*

Theme 4: Encouragement and Motivation Were More Salient Incentives Than Monetary Rewards

Gamification and Rewards: Nonmonetary Rewards

When asked about rewards, most participants expressed that encouragement and nonmonetary rewards were motivating.

> *If you reach a goal, maybe,...you will get in this, a motivating statement, you know, as an incentive or a bunch of balloons goin', confetti exploding, you could maybe get a star.* [Focus group 2, Interviewee 1]
>
> *Maybe even just a motivational somethin', say, some kind of quote...even if it was just, "Good job. You - you did not smoke X number of hours today," or "You smoked less by X percent today." That's the reward right there.* [Focus group 8, Interviewee 3]

A few participants noted that receiving advice or feeling accomplished was rewarding.

> *I just like when [it] lets me know that I've accomplished my goals for today.* [Focus group 7, Interviewee 2]
>
> *I just need somethin' to keep me motivated, like a tip a day.* [Focus group 5, Interviewee 2]

As one participant surmised, the app should be a place, "where you can get encouragement...or you can encourage somebody" [Focus group 2, Interviewee 1].

A few participants expressed that tangible incentives are not as encouraging as the benefits of quitting for one's own health.

> *Yeah, it's not about the money, the incentives. It's what I got to do to live longer.* [Focus group 5, Interviewee 2]
>
> *I have a little girl that I gotta live for, so I'm tryin' to do better with- the smokin'...So that's...motivating me to stop.* [Focus group 6, Interviewee 7]

One participant noted that interacting with others within the app and reaching improvement milestones would be rewarding.

> *I don't really care to be rewarded, because I want to quit smoking. So, I just want the motivation there in the app, tracking...how many cigarettes that I, you know, didn't smoke or whatever, you know, form of tobacco you use. And then just-just having the motivation...* [Focus group 7, Interviewee 8]

Gamification and Rewards: Provision for Diverse Reward Options

Other participants expressed that although money is not the most important motivation, it may be important to some people; hence, allowing participants to choose their incentives was ideal. Some participants mentioned financial rewards, such as gift cards or payments.

> *Some apps give you like daily points if you log in daily. So, if you can get a reward for logging in daily and then it could add up to getting the gift card, getting, you know, PayPal transfers, I think that would be pretty enticing.* [Focus group 7, Interviewee 7]
>
> *It always works better when there's options versus you only have maybe one thing that you can do. So, it might be, oh, a cash app payout or Amazon, but then for the next person they might not really want that, they want something else.* [Focus group 7, Interviewee 4]

Discussion

Principal Findings

As more mHealth apps for tobacco cessation are introduced each year, it is critical that these apps are built to serve everyone, including priority populations. Increasingly, apps tailored for Black adults have been introduced to confront health inequities, with an emphasis on iterative design using formative work from members of the target population [37]. Themes and subthemes from the focus groups revealed a strong preference for certain features, including the type and presentation of content in the app, inclusivity, user experience, tracking capability, professional and social support, gamification, and rewards. Preferred features also aligned with components of the self-determination theory, thereby enhancing the significant role that these features may play in the usage of an mHealth app. These findings provide insight into what app features deter or promote app receptivity, and how to attract Black adults who use tobacco products to use an mHealth app for smoking cessation.

Comparison With Previous Work

In our study, we found that participants desired an app with features that highlight representation and promote inclusivity. To do this, participants recommended that an mHealth app should include testimonials, images, and videos featuring Black adults. Prior studies examining preferred features of mHealth apps for Black women have found that participants desire an app to include a directory of nearby Black female health care providers within their community [38,39]. Having racial concordance with a health care provider can help participants feel they have someone with whom they can relate [39]. One way that representation can be achieved in an mHealth app for tobacco cessation is to have a video testimonial by a Black former tobacco user included in the app, with the person sharing their experience with quitting as well as the health benefits experienced after quitting.

We also found participants desired in-app personalization of content and user experience based on their attributes, such as race, gender, and tobacco use habits and experience. Our findings suggest that to improve the user experience, mHealth apps should allow for personalization based on several factors, not just race. This may lead to a more robust experience for the app user. Further, several participants emphasized how the hypothetical mHealth app should feature a diverse range of people with different perspectives, experiences, and identities. This is consistent with findings in one study on app development where participants desired a diversity of people within an app, specifically wanting "everyone" to be included [40].

Similar to our study, previous research has shown that people are interested in an app with relevant health information and statistics [28]. This finding suggests that when health risks and statistics are personalized, the data are more relatable. Providing personalized content and feedback is associated with increased engagement with mobile app interventions [41]. For example, in an mHealth app for tobacco cessation, including statistics about the rate of cardiovascular disease or cancer for Black tobacco users compared with nontobacco users may increase engagement.

Another important finding in this study is that while participants were in favor of representation within the app, they were understandably hesitant or resistant to an app that focused solely on their racial identity. A previous study on mHealth apps for smoking cessation noted similar concern [28]. This perspective may stem from the justified mistrust of the health community and historical mistreatment of Black people in health care and research [42-44]. In a study that examined mHealth research receptivity among Black men, over a quarter (27%) of men and nearly a third of Black women (30%) identified "mistrust of researchers" as a barrier to participation [45,46].

While a more inclusive but nontargeted app is preferred, outreach strategies to reach the Black community were suggested, including marketing at events for the Black community, marketing by Black influential figures or celebrities on social media, and promotion by Black influencers. This finding suggests that having a known source may engender feelings of trust [47]. Furthermore, in recent years, health campaigns have begun working with influential figures such as pastors or influencers to promote health interventions, including vaccine campaigns [47] and physical activity and healthy eating [48,49]. Influencers can leverage their social influence and reach their audiences directly to promote health interventions [50,51], which could be used to increase app receptivity within the Black community.

In addition, with the high rate of social media use by Black adults [25] and the inclusion of social media as a tool for health promotion and intervention dissemination due to wider reach, scalability, and cost-effectiveness [52], targeted social media campaigns can be beneficial. In tobacco cessation studies, the use of social media platforms for recruitment has shown success, especially for minority populations that are typically hard to reach [49]. In one study by Bricker et al [52], tailoring of a social media campaign to disseminate an mHealth app in a way that focused on self-control and health values was found to be cost-effective and associated with high reach based on the number of clicks and conversion to app installation. The use of social media platforms to disseminate an mHealth app for tobacco cessation targeted for Black adults can increase the likelihood of awareness of the mHealth app, which may convert into more interest and greater use.

Furthermore, prior research to investigate smokers' reactions on social media platforms revealed that comments and the type of reaction such as "love" or "haha" were a reflection of the level of motivation with regard to readiness to quit [53]. This suggests that although social media can be used to enhance the visibility of mHealth apps for tobacco cessation, it can also stimulate motivational processes by serving as a space for cessation-related posts that contain culturally representative narratives, prompting users to share their own motivations and goals for quitting or positively respond to comments that are reflective of their own experiences. This way, social media is not only a tool to promote the mHealth app but can also help to assess qualitative measures such as levels of engagement and motivation to quit.

Our findings also revealed the importance of an in-app community that offers accountability through supportive connections (eg, through health care providers and quit buddy programs). Previous research on mHealth apps has found that many people desire the ability to connect with health care providers within apps [39]. One study examining supportive accountability through counselor monitoring and supportive advice found that having a trusted expert within an mHealth app intervention for smoking cessation led to increased app engagement [54]. In another intervention that used peer-to-peer exchanges between participants seeking smoking cessation resources and communications from experts with treatment-related questions for participants to discuss, doubled sustained abstinence was observed compared to those who only used nicotine patches and had access to a cessation website [55]. Further, the effectiveness of mHealth interventions can be enhanced by human and social support [56,57].

Although open to connections, participants were hesitant and less receptive to having family or social media integrated within the app due to privacy concerns. Other studies examining receptivity to social media on mHealth apps have similarly found that most participants did not want social media integration, citing privacy concerns [39]. Although privacy is a major concern, social media play a significant role by serving as a platform through which health-related behavior change interventions can be delivered and health-related connections or support groups developed in a meaningful way [58,59]. For Black adults who use tobacco products, addressing the privacy concern in a transparent way may improve receptivity to the integration of elements of social media into the intervention.

In contrast, some people prefer having emotional support, such as encouragement, from friends and family [60,61]. Studies have found that stigma against individuals who smoke can prevent them from seeking care [62]. One study found that negative comments from one's social network about their smoking habits can even lead to a decrease in quitting progress

[61]. To combat this concern, any external connections to the app should be optional.

App tracking capability was also an important feature for accountability and has been identified as a desirable feature and method to promote accountability in other studies [39,54]. Several mHealth interventions use the tracking feature to help monitor chronic health conditions and medication intake [63,64]. One qualitative study that explored experiences of Black adults who use tobacco products with a specific mHealth app for smoking cessation identified that the tracking feature was highly preferred [28]. With mHealth apps for tobacco cessation, tracking of the rate of use of tobacco products, cravings, or money saved over a period of time from less use can lead to sustained motivation and cessation-related behavior.

Lastly, we found that while some participants said that financial incentives could attract some people to use the app, many participants did not think financial incentives were necessary to attract users to the app. Prior research suggests that financial incentives can modify or increase adherence to desired behavior changes, like exercise [65], dietary changes [66], or engagement with a state Quitline [67]. However, the sustained provision of financial incentives may not be feasible for mHealth apps long term, nor are they always necessary. In one study examining motivations and barriers to participating in mHealth research among Black males, approximately 40% indicated that offering financial incentives would motivate participation, while only 16% indicated that a lack of financial incentives was a barrier to participation [45]. Similar rates were also seen among Black women, 39% and 15%, respectively [46]. Continued engagement with an app is further augmented by other considerations, like the content shared [68]. Therefore, while financial incentives are desirable, their absence may not significantly deter participation with mHealth interventions. This finding aligns with existing research that the gamification (eg, inclusion of game elements and rewards) of mobile apps can reward and motivate app users [69,70] and promote behavior change [71].

Design Implications for App Development

To enhance the translation and application of our findings to an mHealth app development, we mapped out each subtheme and domain feature into examples of concrete mHealth app features with pertinent processes from the self-determination theory. We also included examples of ethical and privacy considerations, as well as potential metrics to consider during implementation. The map, as outlined in Table 3, shows how representation and inclusivity, personalization, content, accountability, tracking and user experience, social and professional support, and nonmonetary incentives can be operationalized into actual, concrete features within an mHealth app for tobacco cessation.

Table . Design implications for mobile health app with preferred features.

Subthemes and domains of app features	Examples of concrete app features	Self-determination theory components	Privacy/ethical considerations	Potential metrics
Representation and inclusivity	• Video testimonials from other Black tobacco users • Culturally relevant language	• Relatedness • Autonomy	• Avoidance of identifying information of individuals in videos or photos • Opt-in for culturally relevant material, instead of it being the default • Informed consent • Avoidance of racial profiling in content	• Amount of time spent and frequency of viewing culturally relevant information • Rate and frequency of opt-in for culturally relevant material
Personalization	• In-app customization of goals, quit date, reminders, and prompts • Personalized diary	• Autonomy	• User-controlled input of information • Password-protected storage and opt-out option	• Achievement of personal goals • Number and frequency of diary entries
Content	• Health risk statistics relevant to the target user	• Relatedness • Competence	• Ensure credibility of the source of information • Avoidance of the use of sensitive language or medical jargon	• Completion of content sections
Accountability	• Quit buddy, daily check-in	• Autonomy • Relatedness	• User opt-in • Anonymization of profile	• Daily active use
Tracking and user experience	• Tracking tools for mood, cravings, number of cigarettes used, and days without use	• Autonomy • Competence	• User control on the type of information to track or delete • Secure or password-protected storage of tracked data	• Daily active use
Social support	• Anonymous in-app community	• Relatedness	• Anonymization of user profile • User opt-in for participation or for integration with social media	• Number of users who opt-in • Number and frequency of comments and posts on the community board
Professional support	• Quitline or access to a tobacco treatment specialist, trusted expert chat	• Competence	• Informed consent to interact with a professional • Encryption of messages	• Number and frequency of use or messages sent to an expert • Number of times the Quitline button or icon is used
Nonmonetary incentives	• Badges, points	• Competence	• Option for private mode where reward is only seen by the user • No shaming or negative language during relapse	• Rate of badges or points earned

Limitations

A limitation in this study is that findings may not be generalizable. First, we used a convenience sample online and recruited from one social media platform. This may have limited the sample size as well as the diversity of participants in terms of gender, age, educational level, and income. Conducting the focus groups online, as well as the requirement of a history of app usage or willingness to use an mHealth app in the future for participation in this study, further limits the diversity of

perspectives about features of mHealth apps for tobacco cessation. Furthermore, findings may also not be generalizable to Black adults outside of the United States due to cultural differences and other factors that may exist across countries. Obtaining the perception of Black adult tobacco users who are not receptive to the idea of mHealth apps for tobacco cessation may provide additional insight into how best to develop and market these apps in a more inclusive way.

Conclusions

Our findings confirm the need to design mHealth apps for tobacco cessation with features that are reflective of cultural relevance, inclusivity, and diversity of experiences without excessive targeting and support motivation as stated in the self-determination theory [19-21]. Consistent with Perski and colleagues [72], engagement with mHealth apps increases when design features support personal relevance, motivation to quit, and credibility. Our findings provide insight into the improvement of existing apps and future development of an mHealth app that can appeal to Black adults who use tobacco products, potentially increasing app use, engagement, and successful cessation. Findings from this research can also serve as a groundwork for future research to explore the effectiveness of an mHealth app culturally tailored with preferred features optimized to enhance engagement by Black adults who use tobacco products and to identify other factors that may improve usage and lead to better outcomes for the target population.

Funding

This project was supported by the National Institute of Drug Abuse of the National Institutes of Health (NIH) through grant K23DA057416. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

Data Availability

The focus group guide is available in Multimedia Appendix 1. Due to confidentiality and ethical considerations, we are unable to publicly share full transcripts and coded data. Additional nonidentifiable information can be provided upon reasonable request to the corresponding author.

Authors' Contributions

CE, CEK, and AOG contributed to the conceptualization and design of the study. CE and SAC led the focus groups. SAC and RP analyzed the data. CE, SAC, RP, CEK, and AOG contributed to the interpretation of the findings. CE and SAC were involved in the drafting of the manuscript, while all authors were involved in the revision of the manuscript prior to submission. All authors gave approval for the final version for submission.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Focus group guide.
[PDF File, 280 KB - mhealth_v14i1e63340_app1.pdf]

References

1. Henley SJ, Thomas CC, Sharapova SR, et al. Vital signs: disparities in tobacco-related cancer incidence and mortality - United States, 2004-2013. MMWR Morb Mortal Wkly Rep 2016 Nov 11;65(44):1212-1218. [doi: 10.15585/mmwr.mm6544a3] [Medline: 27832048]
2. Tobacco use in the Black American community. Truth Initiative. 2024 Aug 27. URL: https://truthinitiative.org/research-resources/targeted-communities/tobacco-use-african-american-community [accessed 2026-01-23]
3. Alexander LA, Trinidad DR, Sakuma KLK, et al. Why we must continue to investigate menthol's role in the African American smoking paradox.. Nicotine Tob Res 2016 Apr;18 Suppl 1(Suppl 1):S91-101. [doi: 10.1093/ntr/ntv209] [Medline: 26980870]
4. Cancer facts & figures for African American/Black people 2022-2024. : American Cancer Society; 2022 URL: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-african-americans/2022-2024-cff-aa.pdf [accessed 2026-01-23]
5. Health behaviors of adults: United States, 2008-2010. : National Center for Health Statistics; 2013 May Vital and Health Statistics, Series 10, Number 257 URL: https://www.cdc.gov/nchs/data/series/sr_10/sr10_257.pdf [accessed 2026-01-23]
6. Nollen NL, Cox LS, Yu Q, et al. A clinical trial to examine disparities in quitting between African-American and White adult smokers: design, accrual, and baseline characteristics. Contemp Clin Trials 2016 Mar;47:12-21. [doi: 10.1016/j.cct.2015.12.001] [Medline: 26667382]

7. Carroll DM, Cole A. Racial/ethnic group comparisons of quit ratios and prevalences of cessation-related factors among adults who smoke with a quit attempt. Am J Drug Alcohol Abuse 2022 Jan 2;48(1):58-68. [doi: 10.1080/00952990.2021.1977310] [Medline: 34752715]
8. Enyioha C, Hall M, Voisin C, Jonas D. Effectiveness of mobile phone and web-based interventions for diabetes and obesity among African American and Hispanic adults in the United States: systematic review. JMIR Public Health Surveill 2022 Feb 4;8(2):e25890. [doi: 10.2196/25890] [Medline: 35119368]
9. Willis HA, Neblett EW. Developing culturally-adapted mobile mental health interventions: a mixed methods approach. Mhealth 2023;9:1. [doi: 10.21037/mhealth-22-19] [Medline: 36760787]
10. Bhuyan SS, Lu N, Chandak A, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
11. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415. [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
12. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299. [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
13. James DCS, Harville C. Smartphone usage, social media engagement, and willingness to participate in mHealth weight management research among African American women. Health Educ Behav 2018;45(3):315-322. [doi: 10.1177/1090198117714020] [Medline: 28606004]
14. Goodwin RD, Ganz O, Weinberger AH, Smith PH, Wyka K, Delnevo CD. Menthol cigarette use among adults who smoke cigarettes, 2008-2020: rapid growth and widening inequities in the United States. Nicotine Tob Res 2023 Mar 22;25(4):692-698. [doi: 10.1093/ntr/ntac214] [Medline: 36223889]
15. Trinidad DR, Xie B, Fagan P, et al. Disparities in the population distribution of African American and non-Hispanic White smokers along the quitting continuum. Health Educ Behav 2015 Dec;42(6):742-751. [doi: 10.1177/1090198115577376] [Medline: 25794519]
16. Bennett GG, Wolin KY, Robinson EL, Fowler S, Edwards CL. Perceived racial/ethnic harassment and tobacco use among African American young adults. Am J Public Health 2005 Feb;95(2):238-240. [doi: 10.2105/AJPH.2004.037812] [Medline: 15671457]
17. Zhou ES, Revette A, Ritterband LM, et al. Developing a culturally tailored digital health intervention for insomnia in Black women. Transl Behav Med 2024 Feb 7;14(2):117-126. [doi: 10.1093/tbm/ibad056]
18. Wadi NM, Asantewa-Ampaduh S, Rivas C, Goff LM. Culturally tailored lifestyle interventions for the prevention and management of type 2 diabetes in adults of Black African ancestry: a systematic review of tailoring methods and their effectiveness. Public Health Nutr 2022 Feb;25(2):422-436. [doi: 10.1017/S1368980021003682] [Medline: 34435943]
19. 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]
20. Williams GC, McGregor HA, Sharp D, et al. Testing a self-determination theory intervention for motivating tobacco cessation: supporting autonomy and competence in a clinical trial. Health Psychol 2006 Jan;25(1):91-101. [doi: 10.1037/0278-6133.25.1.91] [Medline: 16448302]
21. Choi J, Noh GY, Park DJ. Smoking cessation apps for smartphones: content analysis with the self-determination theory. J Med Internet Res 2014 Feb 12;16(2):e44. [doi: 10.2196/jmir.3061] [Medline: 24521881]
22. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [Medline: 11846344]
23. Ubhi HK, Michie S, Kotz D, Wong WC, West R. A mobile app to aid smoking cessation: preliminary evaluation of SmokeFree28. J Med Internet Res 2015 Jan 16;17(1):e17. [doi: 10.2196/jmir.3479] [Medline: 25596170]
24. 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;41(1):68-73. [doi: 10.3109/00952990.2014.977486] [Medline: 25397860]
25. Social media use in 2021. Pew Research Center. 2021 Apr 7. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2026-01-23]
26. Bricker JB, Copeland W, Mull KE, et al. Single-arm trial of the second version of an acceptance & commitment therapy smartphone application for smoking cessation. Drug Alcohol Depend 2017 Jan 1;170:37-42. [doi: 10.1016/j.drugalcdep.2016.10.029] [Medline: 27870987]
27. Struik LL, Bottorff JL, Baskerville NB, Oliffe J, Crichton S. Comparison of developers’ and end-users’ perspectives about smoking cessation support through the Crush the Crave app. JMIR mHealth uHealth 2019 Mar 7;7(3):e10750. [doi: 10.2196/10750] [Medline: 30843864]
28. Enyioha C, Loufman LM, Grewe ME, et al. Black smokers’ preferences for features of a smoking cessation app: qualitative study. JMIR Form Res 2023 May 30;7:e43603. [doi: 10.2196/43603] [Medline: 37252777]
29. Barroso-Hurtado M, Suárez-Castro D, Martínez-Vispo C, Becoña E, López-Durán A. Smoking cessation apps: a systematic review of format, outcomes, and features. Int J Environ Res Public Health 2021 Nov 6;18(21):11664. [doi: 10.3390/ijerph182111664] [Medline: 34770178]

30. Buller DB, Borland R, Bettinghaus EP, Shane JH, Zimmerman DE. Randomized trial of a smartphone mobile application compared to text messaging to support smoking cessation. Telemed J E Health 2014 Mar;20(3):206-214. [doi: 10.1089/tmj.2013.0169] [Medline: 24350804]
31. Iacoviello BM, Steinerman JR, Klein DB, et al. Clickotine, a personalized smartphone app for smoking cessation: initial evaluation. JMIR mHealth uHealth 2017;5(4):e56. [doi: 10.2196/mhealth.7226]
32. Brown K, Dyas J, Chahal P, Khalil Y, Riaz P, Cummings-Jones J. Discovering the research priorities of people with diabetes in a multicultural community: a focus group study. Br J Gen Pract 2006 Mar;56(524):206-213. [Medline: 16536961]
33. Hayashi P, Abib G, Hoppen N, Wolff LDG. Processual validity in qualitative research in healthcare. Inquiry 2021;58. [doi: 10.1177/00469580211060750] [Medline: 34845941]
34. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
35. Grbich C. Qualitative Data Analysis: An Introduction: SAGE Publications; 2013. [doi: 10.4135/9781529799606]
36. Agaku IT, Alpert HR. Trends in annual sales and current use of cigarettes, cigars, roll-your-own tobacco, pipes, and smokeless tobacco among US adults, 2002-2012. Tob Control 2016 Jul;25(4):451-457. [doi: 10.1136/tobaccocontrol-2014-052125] [Medline: 25899447]
37. Samarasekera U. The rise of racial minority health apps. Lancet Digit Health 2022 Apr;4(4):e218-e219. [doi: 10.1016/S2589-7500(22)00043-7] [Medline: 35337641]
38. McCall T, Ali MO, Yu F, Fontelo P, Khairat S. Development of a mobile app to support self-management of anxiety and depression in African American women: usability study. JMIR Form Res 2021 Aug 17;5(8):e24393. [doi: 10.2196/24393] [Medline: 34133313]
39. Chandler R, Hernandez N, Guillaume D, Grandoit S, Branch-Ellis D, Lightfoot M. A community-engaged approach to creating a mobile HIV prevention app for Black Women: focus group study to determine preferences via prototype demos. JMIR mHealth uHealth 2020 Jul 24;8(7):e18437. [doi: 10.2196/18437] [Medline: 32706723]
40. White MJ, Xie R, Lane H, et al. Organizational trust, usability, and inclusivity are key implementation facilitators for a proposed assets-based mobile health intervention. Transl Behav Med 2023 Jul 1;13(7):465-474. [doi: 10.1093/tbm/ibac108] [Medline: 36999807]
41. Oakley-Girvan I, Yunis R, Longmire M, Ouillon JS. What works best to engage participants in mobile app interventions and e-Health: a scoping review. Telemed J E Health 2022 Jun;28(6):768-780. [doi: 10.1089/tmj.2021.0176] [Medline: 34637651]
42. Dong L, Bogart LM, Gandhi P, et al. A qualitative study of COVID-19 vaccine intentions and mistrust in Black Americans: recommendations for vaccine dissemination and uptake. PLoS ONE 2022;17(5):e0268020. [doi: 10.1371/journal.pone.0268020]
43. Nguyen BT, Brown AL, Jones F, et al. “I’m not going to be a guinea pig”: medical mistrust as a barrier to male contraception for Black American men in Los Angeles, CA. Contraception 2021 Oct;104(4):361-366. [doi: 10.1016/j.contraception.2021.06.001] [Medline: 34118271]
44. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010 Aug;21(3):879-897. [doi: 10.1353/hpu.0.0323] [Medline: 20693733]
45. James DCS, Harville C. Barriers and motivators to participating in mHealth research among African American men. Am J Mens Health 2017 Nov;11(6):1605-1613. [doi: 10.1177/1557988315620276] [Medline: 26634861]
46. James DCS, Harville C II, Whitehead N, Stellefson M, Dodani S, Sears C. Willingness of African American women to participate in e-Health/m-Health research. Telemed J E Health 2016 Mar;22(3):191-197. [doi: 10.1089/tmj.2015.0071] [Medline: 26313323]
47. Moore D, Mansfield LN, Onsomu EO, Caviness-Ashe N. The role of Black pastors in disseminating COVID-19 vaccination information to Black communities in South Carolina. Int J Environ Res Public Health 2022 Jul 22;19(15):8926. [doi: 10.3390/ijerph19158926] [Medline: 35897301]
48. Goodyear VA, Boardley I, Chiou SY, et al. Social media use informing behaviours related to physical activity, diet and quality of life during COVID-19: a mixed methods study. BMC Public Health 2021 Jul 6;21(1):1333. [doi: 10.1186/s12889-021-11398-0] [Medline: 34229651]
49. Pechmann C, Phillips C, Calder D, Prochaska JJ. Facebook recruitment using zip codes to improve diversity in health research: longitudinal observational study. J Med Internet Res 2020 Jun 5;22(6):e17554. [doi: 10.2196/17554] [Medline: 32501274]
50. de Vere Hunt I, Linos E. Social media for public health: framework for social media-based public health campaigns. J Med Internet Res 2022 Dec 14;24(12):e42179. [doi: 10.2196/42179] [Medline: 36515995]
51. Powell J, Pring T. The impact of social media influencers on health outcomes: systematic review. Soc Sci Med 2024 Jan;340:116472. [doi: 10.1016/j.socscimed.2023.116472] [Medline: 38070305]
52. Bricker JB, Santiago-Torres M, Mull KE, Sullivan BM, Mehrotra R. Population-level dissemination of a smoking cessation smartphone app: quasi-experimental comparison of values-based messages in social media advertisements. JMIR mHealth uHealth 2025 Jul 28;13:e71619. [doi: 10.2196/71619] [Medline: 40720905]

53. Watti J, Millner M, Siklósi K, Kiss H, Kelemen O, Pócs D. Smokers' engagement behavior on Facebook: verbalizing and visual expressing the smoking cessation process. Int J Environ Res Public Health 2022 Aug 12;19(16):9983. [doi: 10.3390/ijerph19169983] [Medline: 36011617]
54. Lepore SJ, Collins BN, Killam HW, Barry B. Supportive accountability and mobile app use in a tobacco control intervention targeting low-income minority mothers who smoke: observational study. JMIR mHealth uHealth 2021 Jul 2;9(7):e28175. [doi: 10.2196/28175] [Medline: 34255698]
55. Pechmann C, Delucchi K, Lakon CM, Prochaska JJ. Randomised controlled trial evaluation of Tweet2Quit: a social network quit-smoking intervention. Tob Control 2017 Mar;26(2):188-194. [doi: 10.1136/tobaccocontrol-2015-052768] [Medline: 26928205]
56. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30. [doi: 10.2196/jmir.1602]
57. de Dios MA, Stanton CA, Cano M, Lloyd-Richardson E, Niaura R. The influence of social support on smoking cessation treatment adherence among HIV+ smokers. Nicotine Tob Res 2016 May;18(5):1126-1133. [doi: 10.1093/ntr/ntv144] [Medline: 26116086]
58. Jacobs MA, Cobb CO, Abroms L, Graham AL. Facebook apps for smoking cessation: a review of content and adherence to evidence-based guidelines. J Med Internet Res 2014 Sep 9;16(9):e205. [doi: 10.2196/jmir.3491] [Medline: 25205129]
59. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social media use and access to digital technology in US young adults in 2016. J Med Internet Res 2017;19(6):e196. [doi: 10.2196/jmir.7303]
60. Aschbrenner KA, Naslund JA, Gill L, Bartels SJ, O'Malley AJ, Brunette MF. Preferences for smoking cessation support from family and friends among adults with serious mental illness. Psychiatr Q 2017 Dec;88(4):701-710. [doi: 10.1007/s11126-016-9485-4]
61. Aschbrenner KA, Naslund JA, Gill L, et al. Qualitative analysis of social network influences on quitting smoking among individuals with serious mental illness. J Ment Health 2019 Oct;28(5):475-481. [doi: 10.1080/09638237.2017.1340600] [Medline: 28675331]
62. Madawala S, Enticott J, Sturgiss E, Selamoglu M, Barton C. The impact of smoking status on anticipated stigma and experience of care among smokers and ex-smokers with chronic illness in general practice. Chronic Illn 2023 Sep;19(3):557-570. [doi: 10.1177/17423953221101337] [Medline: 35575240]
63. Donevant SB, Estrada RD, Culley JM, Habing B, Adams SA. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 1;25(10):1407-1418. [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
64. Madujibeya I, Lennie TA, Pelzel J, Moser DK. Patients' experiences using a mobile health app for self-care of heart failure in a real-world setting: qualitative analysis. JMIR Form Res 2023 Aug 15;7:e39525. [doi: 10.2196/39525] [Medline: 37581912]
65. Mitchell MS, Goodman JM, Alter DA, et al. Financial incentives for exercise adherence in adults: systematic review and meta-analysis. Am J Prev Med 2013 Nov;45(5):658-667. [doi: 10.1016/j.amepre.2013.06.017] [Medline: 24139781]
66. Wall J, Mhurchu CN, Blakely T, Rodgers A, Wilton J. Effectiveness of monetary incentives in modifying dietary behavior: a review of randomized, controlled trials. Nutr Rev 2006 Dec;64(12):518-531. [doi: 10.1111/j.1753-4887.2006.tb00185.x] [Medline: 17274494]
67. Bourne DE, Williams R, Osbahr L, Roemhildt M, Villanti AC. Implementation of quitline financial incentives to increase counseling sessions among adults who use menthol tobacco products. Health Promot Pract 2024 Mar;25(2):167-169. [doi: 10.1177/15248399231171143] [Medline: 37118924]
68. Brower J, LaBarge MC, White L, Mitchell MS. Examining responsiveness to an incentive-based mobile health app: longitudinal observational study. J Med Internet Res 2020 Aug 10;22(8):e16797. [doi: 10.2196/16797] [Medline: 32773371]
69. Villalobos-Zúñiga G, Cherubini M. Apps that motivate: a taxonomy of app features based on self-determination theory. Int J Hum Comput Stud 2020 Aug;140:102449. [doi: 10.1016/j.ijhcs.2020.102449]
70. Tran S, Smith L, El-Den S, Carter S. The use of gamification and incentives in mobile health apps to improve medication adherence: scoping review. JMIR mHealth uHealth 2022 Feb 21;10(2):e30671. [doi: 10.2196/30671] [Medline: 35188475]
71. de Oliveira R, Cherubini M, Oliver N. MoviPill: improving medication compliance for elders using a mobile persuasive social game. In: UbiComp '10: Proceedings of the 12th ACM International Conference on Ubiquitous Computing: Association for Computing Machinery; 2010:251-260. [doi: 10.1145/1864349.1864371]
72. Perski O, Blandford A, Ubhi HK, West R, Michie S. Smokers' and drinkers' choice of smartphone applications and expectations of engagement: a think aloud and interview study. BMC Med Inform Decis Mak 2017 Feb 28;17(1):25. [doi: 10.1186/s12911-017-0422-8] [Medline: 28241759]

Abbreviations

mHealth: mobile health

Edited by L Buis; submitted 17.Jun.2024; peer-reviewed by D Pócs, E Ross, M Pai; revised version received 06.Jan.2026; accepted 08.Jan.2026; published 06.Feb.2026.

Please cite as:
Clark SA, Philips R, Kistler CE, Goldstein AO, Enyioha C
Features of Mobile Health Apps for Tobacco Cessation That Appeal to Black Adults Who Use Tobacco Products: Focus Group Study
JMIR Mhealth Uhealth 2026;14:e63340
URL: https://mhealth.jmir.org/2026/1/e63340
doi:10.2196/63340

Failed Implementation of Mobile Access to Electronic Health Records in Home Care: Qualitative Study in Sweden

Lovisa Jäderlund Hagstedt[1,2], MD; Helena Hvitfeldt[1,2], PhD; Maria Hägglund[1,3,4], PhD

[1]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

[2]Tiohundra AB, Stockholm, Sweden

[3]Department of Women's and Children's Health, Uppsala University, Box 256, Uppsala, Sweden

[4]Uppsala University Hospital, Uppsala, Sweden

Corresponding Author:
Maria Hägglund, PhD
Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

Abstract

Background: Digitalization and mobile health (mHealth) technologies hold promise for improving home care delivery. However, many mHealth initiatives fail to achieve their goals. Understanding the reasons behind these failures is critical for informing the successful implementation of mHealth in primary and home care settings.

Objective: This study aimed to explore the implementation process of a tablet computer with an mHealth app providing mobile access to the electronic health record (EHR) in home care, identifying barriers and facilitators to its uptake.

Methods: A tablet with EHR access was introduced at 4 primary care centers and 1 municipal home care organization in Sweden. Participants were nurses and physicians working at the study sites. Focus group discussions and interviews were conducted to obtain a rich understanding of implementation-related issues experienced by the health care professionals. Qualitative content analysis was conducted using the Consolidated Framework for Implementation Research to guide interpretation.

Results: Eighteen health care professionals (16 nurses and 2 physicians) participated in the study. The implementation of the mHealth app was largely unsuccessful. Key barriers included limited functionality of the app, technological immaturity, and unstable infrastructure. Organizational context influenced uptake, especially due to differing EHR systems and varying levels of user engagement. Users who were involved in the development process were more positive, despite the absence of certain functionalities, while those excluded struggled with adoption. Long development and implementation timelines and limited training reduced enthusiasm and negatively affected user engagement. Additional challenges included insufficient implementation planning, lack of leadership engagement, and inadequate resources for support and training.

Conclusions: For mHealth implementations to succeed, tools must meet users' needs and integrate seamlessly with existing eHealth ecosystems and infrastructures. Premature implementations can lead to change fatigue and diminish future engagement. Investments in user-centered design, thorough testing, organizational readiness, and sustained support are essential to realize the potential of mHealth in home care.

(JMIR Mhealth Uhealth 2026;14:e69590) doi:10.2196/69590

KEYWORDS

home care; primary care; work environment; CFIR; mHealth; eHealth; qualitative evaluation; electronic health record

Introduction

As populations age, the need for primary care and home care increases [1], yet resources remain scarce, with serious recruitment challenges concerning educated personnel [2,3]. According to the Swedish Association of Local Authorities and Regions, in the future, such care must be provided closer to the patients. Further, a holistic approach, greater use of digitalization, more collaboration among departments and specialties, and improved patient access are considered necessary. This sentiment is echoed in the World Health Organization's global strategy on digital health, where the purpose of the strategy is to "strengthen health systems through the application of digital health technologies for consumers, health professionals, health care providers and industry towards empowering patients and achieving the vision of health for all" [4]. Electronic health records (EHRs) are used in everyday care throughout Sweden [5]. Most regions use one EHR system throughout hospitals, primary care, and psychiatric care, making patient records available in the entire region. Elder care, however, is the responsibility of municipalities that fall under different legislation [6], and often use separate EHR systems [7-9].

The concept of mobile health (mHealth) has emerged as a subcategory of eHealth and can be defined as medical and public health practice supported by mobile devices [5,10]. The use of smartphones and tablets in health care is increasing. The use of mobile devices for home care in Swedish municipalities has increased significantly in the past few years; in 2020, 41% of the municipalities provided health care staff the possibility to read and write information using a mobile device, compared to 5% in 2015 [11,12]. By 2024, 70% of the municipalities used mobile devices in home care [12].

Further digitalization and the introduction of eHealth and mHealth have been seen as means to optimize health care [4]. Expectations are high that they will be the solution to many of the problems that exist in health care today [13]. However, many projects are unsuccessful [14], and challenges of digitalization have been raised [15]. There are several reasons why implementations of eHealth in health care fail [16,17]. By studying and analyzing why an eHealth or mHealth implementation is unsuccessful, we could learn how successful digitalization in primary and home care can be achieved.

A Cochrane review of mHealth apps showed that the complexities of health care delivery and human interactions defy simplistic conclusions on how health workers will perceive and experience their use of mHealth. Perceptions reflect the interplay between technology, contexts, and human attributes. Detailed descriptions of the program, implementation processes, and contexts, alongside effectiveness studies, will help to unravel this interplay to formulate hypotheses regarding the effectiveness of mHealth [18]. Despite the increasing use of mHealth [19], limited research on mobile EHRs exists, mostly focusing on in-hospital use for bedside EHR access [20,21]. Studies indicate that usability issues of mobile EHRs combined with low digital literacy negatively impact nurses' adoption of the tools [21], and may even have a negative impact on nurses' work environment despite the aim to better support their work [22].

This study was designed to evaluate an mHealth app that provides health care professionals (HCPs) with mobile access to the EHR and describe its impact on work processes and work environment. The project was, however, terminated prematurely since HCP did not adopt and use the system. The collected data were instead analyzed to understand the reasons for the implementation failure.

We aim to understand the reasons for an unsuccessful mHealth implementation and potential success factors by describing lessons learned from the implementation process of a tablet computer with an mHealth app that provides mobile access to EHR in home care.

Methods

Overview

A qualitative approach has been applied to understand the HCPs' expectations before implementation and their experiences of using the system. Individual interviews and focus group interviews have been used to obtain a rich understanding of the implementation-related issues. Due to the limited number of users of the mHealth app, a qualitative approach was deemed suitable to gain an in-depth understanding of the factors causing the failed implementation. The Consolidated Framework for Implementation Research (CFIR) by Damschroder et al [23] was used to guide our analysis.

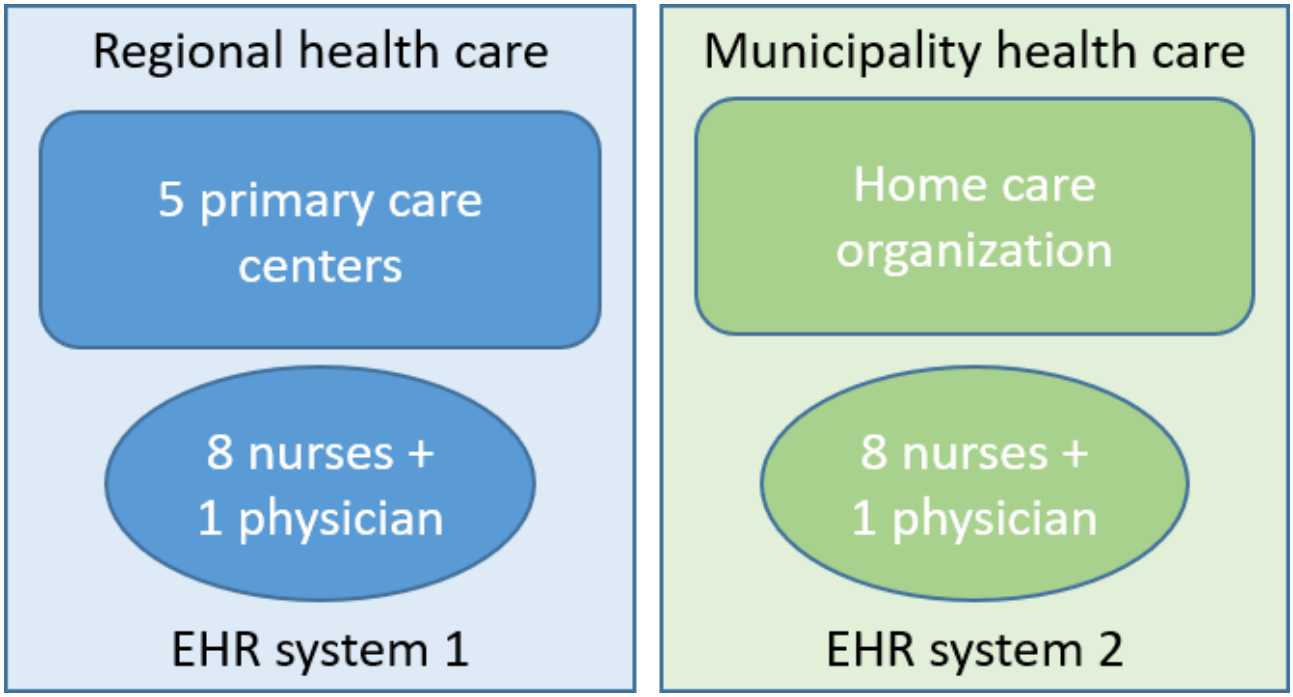
Study Setting

The study was conducted in a public municipality home care organization (MHCO) and the home care sections of 4 public academic primary care centers (PCCs) in a rural part of Region Stockholm.

The home care organization provides medical and supportive care for a diverse range of patients, predominantly older adults but also younger individuals with medical needs. The MHCO has the main responsibility for home care, whereas the PCCs provide some home care, for example, the first 2 weeks after discharge from the hospital and the support and service for persons covered under the Swedish Act Concerning Support and Service for Persons with Certain Functional Impairments. At the time of this study, there was no interaction between the PCCs and the MHCO.

The PCCs all use the same EHR system, accessible through the mHealth app, whereas the MHCO has a different EHR system, lacking accessibility through the mHealth app. Figure 1 presents an overview of the study setting.

Figure 1. Study setting describing the regional and municipality settings and included study participants. EHR: electronic health record.



The Intervention

The intervention consists of a tablet computer with several apps (eg, decision support, online prescription tools, and online supplies ordering), and the new mHealth app that provides mobile access to the EHR (example screenshots from the app are in Multimedia Appendix 1). In the following text, we will refer to the latter as "the mHealth app" and use the term "the mobile tool" when referring to the tablet with all its apps.

The mHealth app on the tablet computer was developed after an initiative by professionals (2 doctors and 2 nurses) at one of the participating academic PCCs. They saw a need for easy access to information, including patient data, and to carry out all parts of their home care work, including documentation, while in the patient's home. The HCP took an active part in an initial need and requirement analysis, and the mHealth app was then developed from their ideas. The first author was involved in some of the app's needs and requirement analysis. Professionals provided feedback on functionality and design but were not further involved in the mHealth app's development. No structured usability assessments were performed.

The project was carried out through a collaboration between academic PCCs and the IT unit of Region Stockholm, along with the private company Chorus, which was responsible for the technical development of the app. As an innovation project, each participating entity covered its own costs. The aim of the pilot project was to evaluate the tool's effectiveness and its potential for broader implementation within the region, with the long-term goal of generating both clinical benefits and economic value. The mobile tool was tested at selected academic PCCs, and the responsibility for approval and rollout was shared between the involved academic PCCs and the IT unit of Region Stockholm.

The mHealth app provided mobile access to the EHR used in regional home care. This included, for example, access to patient record entries, information about the patients to be visited during the day, and the possibility to write patient record entries while in the patient's home, a task otherwise done after returning to the office, as previously described in our study outlining the matching of the functionality of the app to the home care work processes [24]. While the previously published study described the work process in primary care in detail, matching it to the functionality of the study, it did not go into details on the users' experiences and reasons for the failed implementation. Due to technical difficulties, not all parts of the EHR were accessible through the mHealth app. The app did not provide information on patients' medications, referrals, or radiology results because the EHR from which the mHealth app collected the medical data did not support this output. However, it was possible to access information on the patients' previous outpatient and inpatient care visits, diagnoses, and laboratory tests. Thereby, some, but not all the HCPs' needs were met by the mHealth app. The mHealth app was intended to be further developed, and the implementation was therefore initiated.

The tablet computer also provided access to several clinical decision support systems and a mobile web-based prescription system that is used for some of the patients in home care, thus enabling the HCP to manage part of those patients' medications. Thereby, the diverse functions of the tablet computer provided HCPs with a multifaceted mobile tool for home care visits.

There were no costs involved for the users. Use of the mobile tool was strongly encouraged but voluntary.

The mobile tool was first implemented in the home care section of the PCC, which was part of the development of the mHealth app and in one of the other PCCs. In both participating PCCs where the mobile tool was first implemented, some HCPs were avid advocates for the mHealth app.

At a later stage, the mobile tool was implemented in the MHCO and the remaining PCCs. There were no advocates for the mHealth app at either of these workplaces. The mHealth app was further developed later to improve stability and implemented and used in in-hospital care for bedside record access in the Stockholm region. In 2020, Region Stockholm ceased all support of the mHealth app, and it is currently not in commercial operation.

Data Collection

Data for this study were collected through focus groups and semistructured interviews with nurses and doctors working in home care where the mobile tool was introduced. The first author (LJH, female) was responsible for the data collection, with the assistance of 2 Master's students (both female, health informatics students). All 3 had both training in and experience of qualitative data collection, and were supported by the last author (MH, female) with extensive experience in qualitative research. LJH was working as a general practitioner in the region at the time of the study and was well-known to all study participants. An overview of the timeline and data collection is presented in Table 1. Before implementation, a workshop was held where questions were asked regarding the time spent on administrative work related to home visits, experienced stress, use of decision support tools, and the experienced quality of documentation and communication between HCPs.

Table . Overview of the data collection timeline, including the method used, setting, and time for data collection.

Data collection	PCC1[a]	PCC2	PCC3	PCC4	MHCO[b]
Development of an app	2014 - 2015	—[c]	—	—	—
Focus group on expectations	November 2015	May 2016	April 2017	April 2017	April-May 2017
Individual interviews	May 2016	February 2017	February 2018	March 2018	February 2018
Focus group on completion of study	October 2016	—	—	—	—

[a]PCC: primary care center.

[b]MHCO: municipality home care organization.

[c]No data were collected using that method at that specific unit.

The mobile tool was introduced and demonstrated during a workshop where participants had the opportunity to practice using it. The demonstration was conducted by guiding participants through the different functionalities step by step, ensuring a comprehensive understanding of its features and applications. This hands-on approach facilitated engagement and allowed attendees to gain practical experience with the tool.

The original plan was to conduct data collection through (1) preimplementation focus groups, (2) individual interviews after 6 months of use, and (3) a second round of focus group interviews after 12 months of use. The different qualitative methods were chosen to gather a rich dataset; the focus group interviews enabled discussion and exchange of opinions between participants, whereas the interviews allowed more detailed exploration of individual HCPs' experiences of using the mobile tool. We also recognized that individual interviews provide an opportunity for participants to express thoughts and concerns they might be reluctant to share in a focus group setting with other participants present.

Interview guides were developed for individual and focus group interviews (Multimedia Appendices 2 and 3). All interviews and focus groups were recorded and transcribed. Notes were also taken during focus groups and included in the material. All interviews were held in Swedish. The individual interviews lasted 20 - 30 minutes and the focus group interviews lasted 1 - 1.5 hours. Data collection was conducted where most convenient for the participants, that is, at the PCCs or the MHCO's office. Data collection was conducted during working hours. Participant checking of the transcribed interviews was not done to reduce the burden on the study participants.

Preimplementation focus group interviews were held in all the PCCs and the MHCO, where the positive and negative expectations were discussed and reflected upon by HCP who would use the mobile tool.

Individual interviews with nurses and physicians were conducted according to plan to capture the staff's experiences using the mobile tool and its impact on work processes.

Interviews were conducted when the mobile tool had been used for at least 6 months. All interviews were conducted by the first author and the Master's students in the research team. The Master's students conducted interviews in the MHCO. One of the Master's students continued working in the project as a research assistant after graduation and also contributed to the data analysis.

All eligible nurses and physicians working in the setting were invited and accepted to participate in the study. Study participants were approached via email and in-person by the first author through the managers of the MHCO and PCCs. The same participants at each site participated in both the focus groups and interviews.

A second set of focus group interviews was planned at the end of the study, when the mobile tool had been used for 1 year. The PCC involved in the mHealth app's development started testing earlier than the other clinics, and a focus group interview was held a year after they started using the mobile tool to capture their long-term experiences. When the 6-month interviews were conducted at the other clinics, they had stopped using the mobile tool, and the decision was made not to conduct the planned

focus group interviews, as they would not gain further experiences of using the mobile tool.

Data Analysis

All interviews were transcribed verbatim and analyzed using a deductive content analysis approach [25]. The CFIR [23] was used to structure the analysis. The data analysis performed in this study does not include any quantification of responses regarding specific categories.

CFIR identifies 5 domains that are essential to implement an intervention [23]:

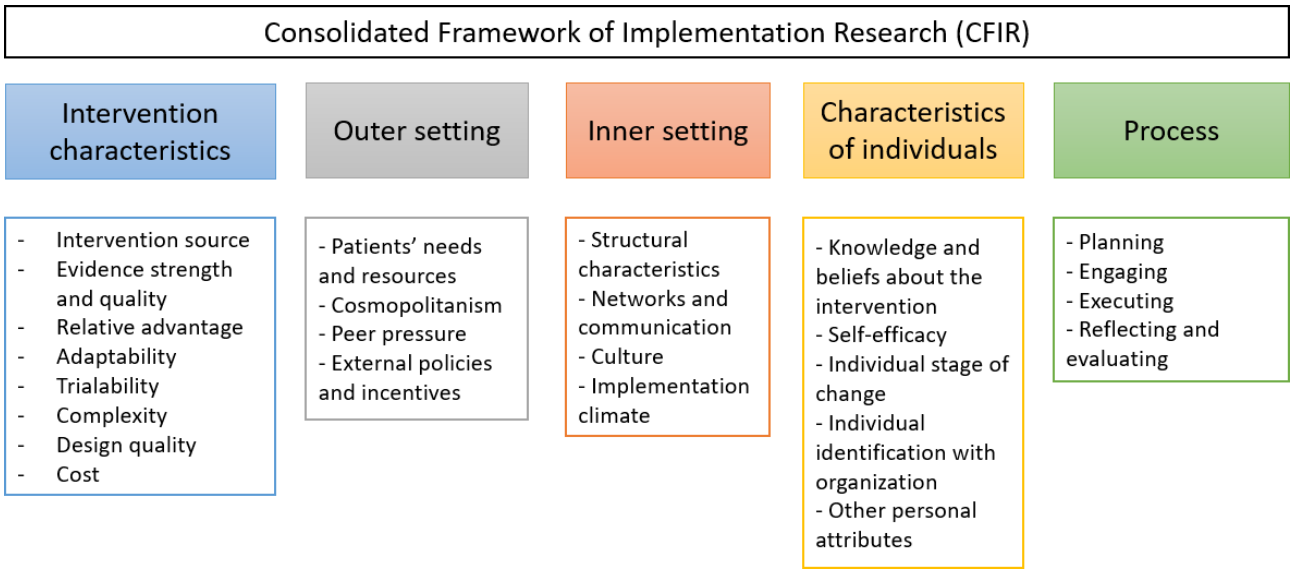
1. Intervention characteristics—features of the intervention itself, such as complexity, evidence strength and quality, and adaptability.
2. Outer setting—the external environment influencing implementation, such as external policies and incentives.
3. Inner setting—organizational factors such as culture, leadership, and available resources within the implementing organization.
4. Characteristics of individuals—traits and actions of the individuals involved in the implementation, including knowledge, beliefs, and self-efficacy.
5. Process—the planning, execution, and reflection processes involved in implementing the intervention.

In CFIR, a domain represents a broad category of factors or determinants that affect the implementation process. A construct, on the other hand, refers to a specific factor within a domain that provides more detailed insight into the aspects of implementation. An overview of the CFIR domains and their constructs is presented in Figure 2.

The 3 authors all actively participated in the analysis. Initial coding was performed by LJH, MH, and one of the Master's students who continued working on the project as a research assistant after graduating. First, all materials were read and reread, and meaning-bearing units were identified. Each coder coded and categorized the meaning-bearing units individually into 1 of the 5 domains of the CFIR framework. HH joined the analysis at this stage, and the 3 authors discussed the coding until a consensus was reached on the interpretation. Where appropriate, relevant CFIR constructs were used to signify our findings, but we did not systematically search for constructs in the material.

Due to the relatively small number of participants, achieving full data saturation was not possible. However, we ensured sufficient information power by including all eligible HCPs across multiple roles and settings. Thematic consistency across discussions suggests key insights were captured despite the limited number of study participants.

Figure 2. An overview of the Consolidated Framework for Implementation Research, including the 5 domains and their constructs.



Ethical Considerations

The Research Ethics Committee at Karolinska Institutet has granted ethical approval for this study (DNR 2015/1457-31/5). All study participants provided written or oral informed consent before participating in the study. In accordance with the Swedish Ethical Review Authorities' requirements, participants were informed that their participation was voluntary and that they could choose to revoke their participation at any time. Transcribed interviews and focus groups were deidentified to protect the participants' privacy, and confidentiality was preserved when reporting results. Study participants did not receive any compensation for their time.

Results

Overview

The 18 participants in the study were all nurses and physicians working in home care in either the home care section of the participating PCCs or the MHCO. A total of 8 nurses and 1 physician from the PCCs (PCC1=2 nurses and 1 physician, PCC2=1 nurse, PCC3=3 nurses, and PCC4=2 nurses), and 8 nurses and 1 physician from the MHCO participated in the study (Figure 1). Participants were predominantly female (17/18, 94.5%), and age ranged between 43 and 64 years. Due to the small number of participants in the study, we chose not to collect

and report further detailed characteristics of the individual study participants to ensure their anonymity.

Using the 5 domains of CFIR, we present our analysis of the HCPs' expectations before implementation and their experiences during the implementation of the mobile tool. We identify the key barriers causing the implementation to fail and the aspects that had a positive influence.

Intervention Characteristics

Before the implementation, HCPs' expectations were high, yet the interviews after 3 and 6 months of using the app revealed that many of the expectations were unmet. Textbox 1 presents an overview of the constructs related to the characteristics of the intervention, with further details from the qualitative analysis presented below.

Textbox 1. Identified categories related to the Consolidated Framework for Implementation Research domain intervention characteristics, divided by expectations before and experiences after implementation.

Expectations before implementation

- Relative advantage
  - Reduction of cognitive workload
  - Time-efficient and timely
  - Real-time access to information
  - Lack of information and functionality
- Complexity
  - Increased complexity and double documentation with many mobile apps
  - Workarounds to adapt to the system may increase cognitive workload
- Design quality
  - Slow and frustrating
  - Immature and unstable

Experiences after use

- Relative advantage
  - Reduction of the cognitive workload
    - Finish work on the site
    - No need for reminders
  - Access to important functions and information
  - Missing functions
  - Reduces unnecessary lab tests and orders
  - Note formatting is worse
- Complexity
  - Adaptations of work processes are required
    - Condensed documentation
    - Preliminary documentation on site
- Design quality
  - Poor design or unclear layout
  - Design improved during the project
  - Immature and unstable

Preimplementation Expectations

HCP had high hopes that the mobile tool would support their work and serious concerns that the technology would not meet their expectations.

Relative Advantage

During the preimplementation focus group interviews, the HCP expressed high expectations that the new mobile tool would help reduce their cognitive workload by allowing them to complete tasks (eg, ordering materials or sending referrals) and finish documentation at the patients' home:

> *To complete the documentation before leaving the patient's home. It would be a dream. You forget so much on the way back to the office.* [Nurse, MHCO]

Real-time access to information would make it possible to make more informed decisions and reduce the need to call colleagues to find information, for example, from the EHR. However, there were also concerns:

> *It's faster to write a post-it note. If it [the mobile tool] doesn't provide everything we need, it may end up in the office* [Nurse, PCC 1]

Staff already knew that some important functionality (eg, the medication list and referrals) would not be available, and concerns were that this would limit the app's usefulness:

> *Frankly, it feels a bit underdeveloped. Several things are missing, like the medication list and x-ray results, which is important when you visit the patient and need to discuss prior events or plans.* [Physician, PCC 1]

Complexity

There were also concerns that missing functionality would result in workarounds and the need for double documentation that might increase the administrative and cognitive workload, especially concerning medications and the lack of necessary templates in the medical record:

> *Above all, I would like access to the medication list. I know I won't, but it would make my job much easier since I organize the patients' pill box dispensers in their homes.* [Nurse, PCC2]

Design Quality

Before the implementation, there were some apprehensions regarding the system's maturity and that it might be slow and frustrating:

> *...that the software is only half-developed and not fully functional. (Physician, PCC 1). Several HCP were worried that the testing would be hampered due to immature technology: If it's malfunctional you won't use it much.* [Nurse, PCC 1]

Intervention Experiences

Overview

The semistructured interviews and the focus group showed that the users' needs and requirements were not sufficiently met by the mobile tool, largely because the mHealth app was not sufficiently adapted to the users' needs. There was, however, a notable difference between how the mobile tool was perceived by the HCP who worked where the intervention was first implemented and who had taken a more active role in the needs analysis and requirements specification, compared to those working where the intervention was later implemented. Those involved in the mHealth app's development and who were part of the initial implementation considered the app much more useful than those who had not been part of the development.

Relative Advantage

One of the nurses responded as follows:

> *Patient visits have become more complete, new questions that arise can be answered quickly.* [Nurse, PCC 1]

The HCP experienced that the mobile tool facilitated work processes and found it fulfilling to be able to answer questions from patients immediately. Another positive effect of the intervention was that it reduced the cognitive workload:

> *I don't need to bring notes that I will later have to remember to document separately on every patient.* [Nurse, PCC 2]

However, not all needs and requirements were met. The concerns about missing functions that were expressed before the implementation were confirmed. Several of the participants stated that the mHealth app needed further development and additional functions:

> *Honestly, it feels undeveloped. Several things are missing, like the medication list and x-ray results, which are important when you're in the patient's home.* [Physician, PPC 1]

Complexity

The limitations in functionality, with missing templates, lack of dictation functionality, and technical challenges with the smaller keyboard on the tablet computer, proved to be a confining factor:

> *The templates are a bit too, well...wrong. So, it becomes very limited, and then it feels easier to take notes by hand.* [Nurse, PCC 1]

Many of the HCPs also felt that their documentation became more concise, and that the on-site documentation in the app was preliminary and needed further development when they returned to the office:

> *The documentation may be a little shorter, sometimes the record entries become shorter.* [Nurse, PCC 1]

For some, the limitations in documentation functionality resulted in them not using that at all:

> *And still, for me, it is faster to dictate than to write. So, I do not write any record entries in the app. I use it to read record entries in the patient's home. Then it works well for the most part.*" [Physician, PCC 1]

Design Quality

The users were not impressed by the app's layout, especially the design of the laboratory results:

> *But you get for example all the hemoglobin results one after the other. It's difficult to read. Seven or eight rows of results for each blood test.* [Nurse, PCC 1]

The users did, however, feel that the design improved somewhat during the project.

Most of the HCPs experienced difficulties with the functionality and deemed the app immature and unstable.

> *Sometimes the internet connection has been bad. Sometimes the application does not work despite an adequate internet connection. It has happened that I have been able to access the medical record, but then it just stops working. I tap and tap but nothing happens. That's it.* [Physician, PCC 1]

Another user stated:

> *All too often it has not worked. Either the internet connection has been inadequate or there is no explanation why I can't log in.* [Nurse, PCC 1]

Outer Setting

In this domain, the expectations of the intervention coincided with the actual experiences of the intervention to a large extent. Textbox 2 shows the categories identified in the outer setting domain.

Textbox 2. Identified categories related to the Consolidated Framework for Implementation Research domain outer setting, divided by expectations before and experiences after implementation.

Expectations before implementation

- Technical infrastructure
  - Previous experiences among the staff with poor internet connectivity in the home care areas caused worry
- Patient needs and resources
  - Health care professionals looked forward to being able to use the mobile device as a tool for learning and patient communication
  - Being able to answer the patients' questions immediately

Experiences after use

- Technical infrastructure
  - Poor internet connectivity
- Patient needs and resources
  - Supports better work processes and improves patient safety.
  - Facilitates patient participation.

Preimplementation Expectations

HCPs were expressing concerns regarding the internet connectivity in the area but were also looking forward to using the mobile tool to meet the needs of their patients.

Technical Infrastructure

Network connectivity is limited in parts of the rural areas of the study setting, and since many of the HCPs had previous experiences with problems due to poor internet connectivity, concerns were expressed before the implementation:

> *The fear is the network connection. That's the thing, and it's hard to solve.* [Nurse, PCC 2]

Patient Needs and Resources

HCPs were hoping that the mobile tool would enable them to improve their communication with patients as well as the patients' participation in their care by using it as both a tool for learning:

> *And you may be able to show patients how to easily access sites for healthcare e-services* [Nurse, PCC 1]

and a means to answer questions asked by the patient:

> *Medical record entries may be useful when you're in the patient's home. Like if a patient asks, ' What did the doctor say at the last visit? I did not understand.'* [Nurse, MHCO]

Intervention Experiences

The HCP's expectations regarding the outer setting were fulfilled to a high degree.

Technical Infrastructure

The limited network connectivity was a real challenge when using the mobile tool. One of the users stated:

> *Our damn bad network connection has proven to be a huge, huge problem.* [Nurse, PCC 2]

Patient Needs and Resources

The staff experienced that the mobile tool supported better work processes and thus improved patient satisfaction:

> *communication with the patient runs smoother because I'm able to answer questions on-site, eg, concerning lab results.* [Nurse, PCC 1]

However, when the mHealth app did not work, it also sometimes rendered other, less positive questions from patients:

> *I've appreciated being able to answer a question that the patient has asked. But sometimes you get questions about the application when it doesn't work. A patient asked me - But if it does not work, why do you use it?* [Physician, PCC 1]

Inner Setting

Overview

All HCPs experienced that the mHealth app was not adapted to their workflow. Therefore, the mobile tool was not seen as entirely useful. This was especially apparent in the MHCO and for the participating physicians since important functions for their workflow were lacking to a larger extent than for the nurses in the PCCs. There were also implications that the implementation process would have benefited from a stronger commitment from the organization and some of the HCPs from the MHCO experienced that there was not enough time for training and preparation before the implementation. Textbox 3 gives an overview of the categories in this domain.

Textbox 3. Identified categories related to the Consolidated Framework for Implementation Research domain inner setting, divided by expectations before and experiences after implementation.

Expectations before implementation

- Structural characteristics
  - Concerns that there would be no time to complete the additional tasks required to manage the system
  - Concerns that the lack of home care visits would make the system less useful and rarely used

Experiences after use

- Structural characteristics
  - With more home care visits, the need for the system increased
- Networks and communications
  - Unclear responsibility and support functions
- Implementation climate
  - Same job, new technology

Preimplementation Expectations

Overview

There were no mentions of the inner setting during the preimplementation workshops, and predominantly concerns were raised.

Structural Characteristics

Staff from the PCCs expressed concerns that the app might not be very useful and that there would be few opportunities for usage due to the few home care visits in primary care. There were also concerns that the management of the app would create additional tasks that there would not be any time for:

> *We might not have time for the increased administration.* [Nurse, MHCO]

Intervention Experiences

The concerns were mostly addressed, but issues related to the support within the organization were also experienced.

Structural Characteristics

Staff who had used the mobile tool for some time reflected that it would be more useful with additional home care visits:

> *If I was to have homecare visits a whole day, day after day, then I think it would be good to be able to do all the documentation immediately. Then I think it really would improve patient safety.* [Nurse, PCC1]

Networks and Communications

The responsibility for technical support of the mobile tool and mHealth app was unclear to many of the participating HCP:

> *But I don't know where to turn for help when it doesn't work. Therefore, I have unfortunately not used it for more than a month.* [Nurse, PCC2]

Another person stated:

> *Well, it's like; who can you ask?!* [Nurse, MHCO]

Implementation Climate

In many ways, the implementation of the mobile tool did not change the HCP's perceptions of their work:

> *It's simply a new technology for the same job.* [Physician, PCC 1]

Characteristics of Individuals

As previously described, the user's tolerance for challenges differed. Some continued using the mobile tool despite difficulties, whereas others stopped using it. The categories related to this domain are presented in Textbox 4.

Textbox 4. Identified categories related to the Consolidated Framework for Implementation Research domain characteristics of individuals, divided by expectations before and experiences after implementation.

Expectations before implementation

- Knowledge and beliefs about the intervention
  - Concerns that the technology will steal time and focus from the patient
  - Concerns of becoming too dependent on the technology
  - Concerns that the project will amount to nothing
- Self-efficacy
  - To dare to try the new technology
- Individual stage of change
  - Looking forward to getting to test
- Individual identification with organization
  - Worry about personal responsibility for the devices

Experiences after use

- Knowledge and beliefs about the intervention
  - Positive that this type of intervention could provide benefit
  - Unreliable technology has required patience and persistence
  - Needs further development and additional functions
- Self-efficacy
  - Embarrassing when technology does not work
  - Improved work processes improve patient safety, credibility, and work satisfaction
  - Confidence in using mobile technology
- Individual stage of change
  - Not used to this way of working
  - Creates a need for new knowledge
  - Have not used the technology

Preimplementation Expectations

Before the implementation began, users expressed eagerness to test the tools they had been waiting for, despite the already described concerns relating to, for example, internet connectivity and mHealth functionality.

Knowledge and Beliefs About the Intervention

HCPs expressed concerns about whether the mobile tool would influence their relationship with the patients, worrying that the technology would shift time and focus from the patient to the technology. They expressed that "there might be too much focus on technology during the visits" (physician, PCC1), and that there could be a "risk for less contact with the patient if the tablet takes more time" (nurse, PCC4).

Another concern expressed was the risk of becoming too dependent on the technology:

> *To become dependent on the tablet, and then when it breaks down you don't know what to do.* [Nurse, MHCO]

Participants also expressed concerns that the technology might never be ready for use or that its functionality would be reduced to the point that the system would become unusable.

Self-Efficacy

Before implementation, the HCPs' expectations varied. Some were confident in their ability to use it, while others were a bit more hesitant and stressed that daring to try would be important:

> *One way is to just start using the tablet; perhaps it'll go better than we think…* [Nurse, MHCO]

Individual Stage of Change

Similarly, individuals' readiness to try the technology varied, but most were eager to start testing the mobile tools.

> *To dare to start using it and see what happens. I think there are several areas of use. You can find all sorts of online services [mentions Swedish e-health services such as the national patient portal, portals for ordering materials, and online prescription tools]. And perhaps you can help the patient to find the patient portal [1177.se] by using the tablet.* [Nurse, PCC 3]

Those involved from the start were more positive and eager to test the mobile tools, having waited a long time for the pilot to start, whereas the others were more hesitant.

Individual Identification With the Organization

Concerns regarding personal responsibility for the device were only raised during the preimplementation focus groups. Participants had concerns that they might forget the tablet at the patient's home, drop it, and break the screen. They worried this could be either a waste of resources for the organization or that they would personally be responsible for the costs.

Intervention Experiences

Although the participants' characteristics did not necessarily change during the implementation, their self-efficacy and beliefs about the intervention appear to have been affected.

Knowledge and Beliefs About the Intervention

The study participants remained positive that the intervention could provide benefits, for example, by making work processes more time efficient and improving patient safety. One of the participants who had used the tool the most said:

> *I don't have to prepare, which usually takes 5 minutes, to print eg, the list of medications, and the latest lab results* [Physician, PCC1]

Care planning in the patients' home was also experienced as more efficient:

> *I can finish it at the patient's home, with the patient. You can discuss what to do next, and then that's done.* [Physician, PCC 1]

To complete the work while in the field was experienced as reducing the risk of forgetting tasks or losing information:

> *[...] and for the patients and the staff [at the nursing home], that they don't have to remind me, 'did you remember that? now this hasn't arrived, why not?,' 'Oh, I didn't order that, I forgot, or I didn't see the note….' It feels more organized.* [Nurse, PCC 1]

However, dealing with unreliable technology required patience and persistence from HCP.

> *The biggest challenge has been not to lose your patience. It's like well, now it doesn't work. I must try the next time again, and the time after that. And that has been difficult since I usually only do this kind of work once a week, so it has been a challenge to not just give up but to instead try and try again.* [Nurse, PCC 1]

Only the most enthusiastic users had the patience to continue trying despite the unstable technology. The other users, especially those who worked in the MHCO, gave up on the first hurdle and never used the mobile tool.

Self-Efficacy

Unstable technology also affects the HCP's sense of professionalism and could be a cause for embarrassment in front of patients:

> *It feels embarrassing when it doesn't work when you use the tablet. I can live with that, but it's a little embarrassing. Then it would've been better not to have it [the mobile tool]* [Physician, PCC 1]

One of the HCPs had problems using the mobile tool in the presence of the patient's relatives:

> *That makes me stressed because the technology is new to me. When five people ask me questions at once. I haven't been able to handle that. [regarding not finding the necessary information on the mobile tool].* [Physician, PCC 1]

Those who did use the mobile tool found that it improved work processes, patient safety, credibility, and work satisfaction.

> *What's been very positive is that I can finish my work in the patient's home. I know that I won't forget anything. The patient won't be left without medications, which is good. Because it has been stressful before, I haven't always had the time to finish the work immediately when I'd returned to the workplace… so, I find it to be satisfactory as well, to be able to finish the job and not have any loose threads.* [Nurse, PCC 1]

They also felt confident using the mobile technology:

> *It's been good. It's easy to connect and I think it works well. I've not experienced any challenges* [Nurse, PCC 2]

However, there was a significant discrepancy between the two contexts, where the HCPs in the PCCs were much more positive. In the MHCO, most of the HCPs did not use the mobile tool at all.

Individual Stage of Change

As mentioned earlier, not all HCPs used the mobile tool:

> *I may have a maximum of five [patients]. And then I might as well remember that information or make some notes until I dictate [the record entries]* [Physician, PCC 1]

One reason was that the users were not used to this way of working. In some cases, the users did not like how record entries were presented:

> *I could edit the record entries later if I only write a few keywords [while in the patient's home]. But I don't know, I don't find the entries aesthetically*

> *appealing. I think they look much better when I do it my way.* [Nurse, PCC 2]

In other cases, they had no prior experience writing directly in the medical record, being used to dictating all record entries:

> *Well, I don't have that much experience. I know how to use the iPad, but it feels like I only use it to read record entries, not to write them myself. I usually don't [write record entries] so I don't do it here either.* [Physician, PCC 1]

The users believed there were other possible areas where the mobile tool could be useful:

> *I think it's more that you don't have enough imagination. You need to find other examples of what you can use it for. You just must use it more and discover even better ways of using it.* [Nurse, PCC1]

It was believed that it would be better to use the mobile tool regularly to reach its full potential:

> *If I had to use this every day, then maybe I would become more comfortable with it* [Nurse, PCC 1]

Process

The implementation process was not addressed in interviews with HCP, as this was not the focus of the original project. We will reflect on the implementation process in the discussion.

Discussion

Principal Findings

The implementation of the mHealth app was unsuccessful, despite reports of positive experiences and potential benefits. The main reason for the failed implementation was that the app did not sufficiently support the users' needs. Immature technology and unstable technical infrastructure also contributed. The findings from this implementation study highlight the critical importance of aligning digital health interventions with user needs and existing infrastructure, particularly in the context of aging populations and resource-constrained health care systems. For digitalization to be effective and sustainable, strategic approaches must prioritize usability, cross-organizational compatibility, and user involvement throughout development and implementation. Moreover, adequate resourcing—including time, training, and leadership engagement—is essential to prevent change fatigue and to ensure long-term adoption and impact of mHealth technologies.

Before implementation, HCPs had high expectations that the app would facilitate work processes and reduce cognitive workload by allowing real-time access to information. There were also concerns, mainly due to prior experiences of poor internet connectivity and immature and unstable technology, but also that the app might increase the workload, shift time, and focus from the patient, and that the project would amount to nothing. Ultimately, both expectations and concerns were confirmed. Another preimplementation concern expressed by some HCPs was that they would become too dependent on the new technology. This was not brought up again postimplementation, and we interpret this as being due to the failed implementation. The concern could have negatively influenced the HCP's willingness to adopt the technology, but likely this was not the most important factor for the failed implementation.

Some users, particularly those involved in the mHealth app's development, found value in the app despite its shortcomings and kept trying time after time, undeterred by the immature technology. However, users who worked in the MHCO and had not been involved in the design process gave up at the first hurdle and did not use the mobile tool. Data on usage of the mHealth app, for example, the number of log-ins per day or number of active users, were not available during the implementation process. Access to such data could potentially have guided the implementation process so that action could have been taken when usage was low [26].

End user involvement in the design process, often referred to as participatory design [27-29], has proven to be essential in both ensuring that the software meets the end user's needs and supports their work processes, and increases acceptance among the users [30,31]. In addition to the difference in participation during the development process, the MHCO also used another EHR system in their daily work that was not integrated with the mobile system. The benefit of being able to at least begin documentation in the patient's home was lost to them, further lowering their incentives to engage with the system. Lack of interoperability and integration between eHealth systems has previously been identified as an important barrier to successful implementation [16]. Physicians who made few home visits and used dictation for EHR documentation in everyday work found the mHealth app less useful. Again, the importance of capturing the perspectives of all potential end users in the design process was highlighted [31,32].

There are several lessons to be learned from this. First, that outcome depends on context, and second, that user engagement in both the development and the implementation processes is crucial for a successful implementation. In our study, both lack of user involvement [30,33] and poor interoperability [7,17] hurt the acceptability of the system, contributing to a system that did not support the user's work processes [24]. The importance of an interoperable digital health ecosystem is stressed in the World Health Organization digital health strategy, defining it as a "digital interoperable information technology infrastructure that is primarily used by the health care community across all care settings, in particular by health care providers, health service providers and patients" [4]. With such an infrastructure in place, the integration and implementation of the mHealth app would likely have been more successful. The lengthy development and implementation process also affected user engagement negatively. Studies of user involvement's relationship with user satisfaction have indicated that user involvement can overcome frustration from delays and prolonged projects [30]. In our study, we did see higher user satisfaction and tolerance among users who had been more actively involved in the design process, but it was not enough to overcome the implementation challenges completely. Additionally, for less engaged staff, the barrier to start using a new technology was high, and they also had a greater need for additional support, whereas users with high engagement, often

referred to as “champions” in the implementation science literature, kept trying regardless and often found new areas of use for the technology. The most common reason for discarding an eHealth service is issues with the technology [19,34]. eHealth services often are premature and thus do not adequately support the work in home care [19]. Technical issues [19] include everything from a lack of suitable infrastructure to usability issues, the latter referring to whether the product provides the right utility and the degree to which it helps users reach their goals in an effective, efficient, and safe manner, while still being easy to learn and remember how to use [35]. Usability of the mHealth app should have been evaluated more extensively before the implementation, for example, through user-driven scenario-based usability testing [36], or expert- or user-driven heuristic evaluations [37,38].

In our case, considerable problems with internet connectivity, immature technology, and a lack of functionality contributed to the low app usage.

Other studies and reviews confirm the challenges described by the users in this study. User involvement, interoperability, reliability of connection, and technology and infrastructure can be considered facilitators if handled appropriately in the implementation of eHealth, or barriers if they are not considered [12,17]. In a review of 43 studies of mHealth implementation experiences from primary health care services, the main results were that health workers appreciated mHealth when it improved feedback, speed, and workflow [17]. Challenges stated were that mHealth sometimes created more work, resulting in some preferring paper instead. Some health workers found the decision support opportunities useful, and others thought it threatened their clinical skills [18]. The aspect of workflow was found substantial for the success or failure of eHealth interventions particularly if workload increased, if workflow was interrupted, and if there was an alignment with clinical processes [17].

CFIR Suitability for mHealth Implementation

CFIR is developed to study implementation processes, not specifically new technology. We chose to include “technical infrastructure” as a new construct in the domain “Outer setting” since the effect of new technology depends on its functionality. A more specific focus on the usability of the eHealth or mHealth intervention could also be included under intervention characteristics, but we chose to use the “design quality” construct here.

The implementation process is crucial for the success of an implementation and for a successful implementation, all stages of the implementation must be properly executed. Our study was not originally designed to study the implementation process, and subsequently, material on some aspects of the implementation process is missing since no questions regarding those aspects were asked in the interviews. We see, however, that the implementation of the mobile tool most likely would have benefited from better planning, execution, support functions, and resources. When implementing complex interventions, feasibility and pilot studies are essential [35,39]. This is especially important when implementing mHealth where lack of usability is often the cause of failed implementations. This could be further specified in the CFIR domain “process” when studying new technology. A specific challenge of implementing the mHealth app in a rural home care area was the physical distance between units and the challenges of keeping day-to-day contact with and providing support to the HCP. The need for a specific facilitator role, that is, a person dedicated to supporting the implementation process and ensuring problems are identified and addressed, was apparent, as is also stressed in other implementation theories, for example, the Promoting Action on Research Implementation in Health Services (PARIHS) framework [40]. In line with the integrated-Promoting Action on Research Implementation in Health Services (i-PARIHS) framework and other implementation models, the facilitator role typically requires a combination of technical, organizational, and interpersonal competencies. Core competencies include basic technical troubleshooting and familiarity with the intervention, skills in user training and coaching, and the ability to communicate effectively across different organizational levels, including leadership. Responsibilities often span the entire implementation process: supporting planning and readiness activities before rollout, providing hands-on assistance and problem-solving during implementation, and ensuring ongoing support and feedback loops during maintenance. While the exact scope and emphasis of these tasks must be adapted to the local context and resources, these elements are commonly highlighted as critical for successful implementation.

There is a strong belief in the potential for eHealth and the introduction of new technology to revolutionize health care and that it will be necessary for health care and primary care in the future [4]. The introduction of new technology, however, often fails [17,41], and there is a risk that premature implementation of new technology will exhaust HCP for future projects and implementation. The challenges and concerns must be considered to move forward successfully [15]. According to the CFIR [23,42] and i-PARIHS [40] frameworks, the individuals’ previous experiences and beliefs about the intervention will impact the implementation’s success. When HCPs are exposed to immature digital solutions with poor usability, we may risk a negative impact on future implementation and innovation projects. To avoid this, we conclude that it is important to ensure that new technology has the necessary functionality and stability before implementation and that there are routines in place in the organization to guide decision-making and implementation processes. Structured contingency planning for technical failures, such as providing clear fallback procedures, timely support, and communication strategies, could sustain user engagement by reducing frustration, maintaining workflow continuity, and reinforcing trust in the implementation process despite technological setbacks.

Strengths and Limitations

The study was guided by the CFIR framework, which is a well-established model for understanding implementation processes [23]. This adds theoretical rigor and structure to the analysis, ensuring that a broad range of influential factors is considered. In our analysis, we used the original CFIR framework which was available when we began our analysis. More recent versions of the framework [42] have added new constructs; however, the domains remain the same, and we

believe that using the most recent version of CFIR would not have substantially influenced our results.

The use of interviews (both individual and focus group) allowed for an in-depth exploration of the HCPs' experiences and perspectives. This is particularly useful given the small number of users of the mHealth app, making qualitative methods appropriate for capturing rich, detailed insights. The study also collected data from both nurses and physicians across different settings (PCCs and MHCO), which ensured a variety of perspectives were included in the analysis. The small sample size (8 nurses and 1 physician from each of the two groups) may limit the generalizability of the findings to other settings or health care systems. Additionally, the majority of participants were female, which may impact the diversity of perspectives, yet it reflects the current workforce in Swedish home care. To maintain anonymity, the study does not report detailed demographic information about participants, such as years of experience, which could be useful for understanding how these factors might influence perceptions of the mHealth app. We also do not include any quantification of responses, which means that while rich qualitative data were captured, the relative importance of different factors or themes was not assessed. This could limit the ability to weigh certain barriers or facilitators more heavily than others; however, considering the qualitative nature of our dataset, we argue that this is the best approach as quantifying the results may be rather misleading. Using a validated instrument to measure health care professionals' self-efficacy could be a way to further understand how self-efficacy impacts adoption and abandonment, and how it may change over time during an implementation project. This should be explored further in future research.

A strength of the study is that the data collection spanned several years, and follow-up interviews after 6 months of using the mobile tool captured longer-term experiences and effects, offering insights into sustained usage and challenges. However, the study took place from 2015 to 2017, with final data collection in 2018, which could render the results of the study out of date. Yet the findings remain highly relevant in today's digital health landscape. Despite technological advancements, the fundamental implementation challenges we identified, such as poor alignment with clinical workflows, limited user engagement, inadequate training and support, and insufficient organizational readiness, persist in contemporary mHealth and eHealth initiatives. Implementation projects continue to experience implementation challenges across diverse settings, also in large-scale initiatives. For example, recent EHR rollouts in Sweden [43], Norway [44], and Finland and Denmark [45] encountered major disruptions due to user dissatisfaction, lack of preparation, and technological shortcomings. These cases underscore the enduring importance of addressing the sociotechnical and organizational dimensions of digital health implementations.

The first author was involved in the needs and requirement analysis for the mHealth app and also conducted the interviews. This dual role might introduce potential bias into the data collection and analysis process, as prior involvement may influence how the data are interpreted or what aspects are emphasized. Throughout the analysis process, we have been vigilant about this, and the involvement of Master's students with no prior connection to the project facilitated an unbiased data collection. The first author's familiarity with the context and the app has also been an enabling factor, to both data collection and by facilitating insights into both the design and implementation processes that would otherwise have been challenging to obtain in retrospect.

These strengths and limitations provide important context for interpreting the findings of the study and considering its applicability to broader implementation efforts.

Conclusions

We conclude that new technology must be stable and have the desired functionality before implementation for success. The functionality of a new technology must support the users' needs and be sufficiently integrated with other IT systems in the health care organization. How the implementation process is executed is important, but the implementation will not be successful if the new technology does not properly support the users' needs. In the case presented here, implementation failed because the technology was unstable and thus unreliable, not sufficiently supportive of the users' needs, and not adapted to the users' workflow. There were, however, positive experiences too, and mHealth still has a strong potential to support HCPs working in home care. To achieve this potential, we must ensure that sufficient resources are allocated to both design and development, evaluation and feasibility studies, and support and engagement during the implementation of mHealth in the future.

Acknowledgments

We thank Gabor Revay and Richard Scholvin at the Department of eHealth and IT, Region Stockholm, Health Care Service, Sweden, for their support and work on the project. We also thank the Master's students Sofia Rosborg (who continued working as a research assistant in the project) and Malin Lindberg for their contributions to the project.

We used generative artificial intelligence technology Grammarly (Grammarly Inc) to enhance the grammatical accuracy and vocabulary richness.

Funding

This work received funding from Innovationsfonden (20151143), FORTE – the Swedish Research Council for Health, Working Life and Welfare supports "MobEVAL" (2017- - 01962), and AFA Försäkring "ePrIm" (190210).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the qualitative nature of the materials and the terms of the ethical approval, but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MH, LJH
Data curation: MH, LJH
Formal analysis: LJH, MH, HH
Funding acquisition: MH, LJH, HH
Investigation: LJH
Methodology: MH, HH, LJH
Project administration: MH
Supervision: MH, HH
Visualization: MH
Writing – original draft: MH, LJH, HH
Writing – review & editing: LJH, MH, HH

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the mobile health app.
[DOCX File, 998 KB - mhealth_v14i1e69590_app1.docx]

Multimedia Appendix 2
Interview guide focus group interview—expectations of the intervention.
[DOCX File, 13 KB - mhealth_v14i1e69590_app2.docx]

Multimedia Appendix 3
Interview guide—evaluation after the intervention.
[DOCX File, 14 KB - mhealth_v14i1e69590_app3.docx]

References

1. Lanzieri G. Long-term population projections at national level. Eurostat Stat Focus 2006 Feb 28 [FREE Full text]
2. Russo G, Perelman J, Zapata T, Šantrić-Milićević M. The layered crisis of the primary care medical workforce in the European region: what evidence do we need to identify causes and solutions? Hum Resour Health 2023 Jul 14;21(1):55. [doi: 10.1186/s12960-023-00842-4] [Medline: 37443059]
3. Health and care workforce in Europe: time to act. : World Health Organization; 2022 URL: https://iris.who.int/handle/10665/362379 [accessed 2025-03-02]
4. Global strategy on digital health 2020-2025. World Health Organization (WHO). 2021. URL: https://www.who.int/health-topics/digital-health#tab=tab_1 [accessed 2025-12-05]
5. Rapport ehälsa och IT i landstingen 2018 - eHälsa och IT i landstingen maj 2018 Inventering på - Studocu [Website in Swedish]. URL: https://www.studocu.com/sv/document/hogskolan-dalarna/arytmi-kurs/rapport-ehalsa-och-it-i-landstingen-2018/15115952 [accessed 2025-03-02]
6. Lind L, Sundvall E, Karlsson D, Shahsavar N, Ahlfeldt H. Requirements and prototyping of a home health care application based on emerging JAVA technology. Int J Med Inform 2002 Dec 18;68(1-3):129-139. [doi: 10.1016/s1386-5056(02)00071-0] [Medline: 12467797]
7. Hägglund M, Scandurra I, Koch S. Studying intersection points - an analysis of information needs for shared homecare of elderly patients. J Inf Technol Healthc 2009;7(1):23-42 [FREE Full text]
8. Koch S, Hägglund M, Scandurra I, Information Resources Management Association (IRMA) (ed.). Informatics and socio-technical challenges when designing solutions for integrated ecare. In: Healthcare Ethics and Training: Concepts, Methodologies, Tools, and Applications: IGI Global Scientific Publishing; 2017. [doi: 10.4018/978-1-5225-2237-9.ch011]
9. Hägglund M, Scandurra I, Koch S. Supporting citizen-centered care for seniors - experiences from two swedish research projects. Presented at: 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS); Jun 20-22, 2012. [doi: 10.1109/CBMS.2012.6266377]
10. MHealth: new horizons for health through mobile technologies: second global survey on ehealth. World Health Organization. URL: https://iris.who.int/handle/10665/44607 [accessed 2025-03-02]

11. E-hälsa och välfärdsteknik i kommunerna 2017 [Report in Swedish]. : Socialstyrelsen; 2017 URL: https://www.socialstyrelsen.se/contentassets/3bdc30a5fa3f4adbb4f3e279848cbc19/2017-4-22.pdf [accessed 2025-11-15]
12. E-hälsa och välfärdsteknik i kommunerna 2024 [Report in Swedish]. : Socialstyrelsen; 2024 URL: https://www.socialstyrelsen.se/contentassets/4c3331fe414a4749a3e57790aab363b5/2024-5-9099.pdf [accessed 2025-11-15]
13. van der Kleij R, Kasteleyn MJ, Meijer E, et al. SERIES: eHealth in primary care. Part 1: Concepts, conditions and challenges. Eur J Gen Pract 2019 Oct;25(4):179-189. [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
14. Vision e-hälsa 2025 [Report in Swedish]. : Regeringskansliet (Socialdepartementet) and Sveriges Kommuner och Landsting; 2025 URL: https://ehalsa2025.se/wp-content/uploads/2021/02/vision-e-halsa-2025-overenskommelse.pdf [accessed 2025-12-20]
15. Öberg U, Orre CJ, Isaksson U, Schimmer R, Larsson H, Hörnsten Å. Swedish primary healthcare nurses' perceptions of using digital eHealth services in support of patient self-management. Scand J Caring Sci 2018 Jun;32(2):961-970. [doi: 10.1111/scs.12534] [Medline: 28960451]
16. Lennon MR, Bouamrane MM, Devlin AM, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a National Digital Health Innovation Program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42. [doi: 10.2196/jmir.6900] [Medline: 28209558]
17. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235. [doi: 10.2196/10235] [Medline: 29716883]
18. Odendaal WA, Anstey Watkins J, Leon N, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942. [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
19. Rydenfält C, Persson J, Erlingsdottir G, Johansson G. eHealth services in the near and distant future in Swedish home care nursing. Comput Inform Nurs 2019 Jul;37(7):366-372. [doi: 10.1097/CIN.0000000000000536] [Medline: 31135467]
20. Thompson BW. The transforming effect of handheld computers on nursing practice. Nurs Adm Q 2005;29(4):308-314. [doi: 10.1097/00006216-200510000-00004] [Medline: 16260994]
21. Hsu SC, Liu CF, Weng RH, Chen CJ. Factors influencing nurses' intentions toward the use of mobile electronic medical records. Comput Inform Nurs 2013 Mar;31(3):124-132. [doi: 10.1097/NXN.0b013e318270100b] [Medline: 23114391]
22. Heponiemi T, Kaihlanen AM, Gluschkoff K, et al. The association between using a mobile version of an electronic health record and the well-being of nurses: cross-sectional survey study. JMIR Med Inform 2021 Jul 6;9(7):e28729. [doi: 10.2196/28729] [Medline: 34255704]
23. 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. Implementation Sci 2009 Dec;4(1):50. [doi: 10.1186/1748-5908-4-50]
24. Rosborg S, Lindberg M, Ramukumba M, Jäderlund Hagstedt L, Hägglund M. Exploring mHealths fit to workflow in homecare - a case study in Sweden. Stud Health Technol Inform 2019 Aug 9;265:54-59. [doi: 10.3233/SHTI190137] [Medline: 31431577]
25. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
26. Haugen HA. Beyond Implementation: A Prescription for Lasting EMR Adoption: Magnusson Skor; 2010.
27. Bødker S, Iversen OS. Staging a professional participatory design practice: moving PD beyond the initial fascination of user involvement. 2002 Presented at: NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction; Oct 19-23, 2002; Aarhus Denmark p. 11-18. [doi: 10.1145/572020.572023]
28. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
29. Bano M, Zowghi D, da Rimini F. User satisfaction and system success: an empirical exploration of user involvement in software development. Empir Software Eng 2017 Oct;22(5):2339-2372. [doi: 10.1007/s10664-016-9465-1]
30. Berg M. Implementing information systems in health care organizations: myths and challenges. Int J Med Inform 2001 Dec;64(2-3):143-156. [doi: 10.1016/s1386-5056(01)00200-3] [Medline: 11734382]
31. Scandurra I, Hägglund M, Koch S. From user needs to system specifications: multi-disciplinary thematic seminars as a collaborative design method for development of health information systems. J Biomed Inform 2008 Aug;41(4):557-569. [doi: 10.1016/j.jbi.2008.01.012] [Medline: 18394969]
32. Scandurra I, Hägglund M, Koch S. Application of the multi-disciplinary thematic seminar method in two homecare cases - a comparative study. Stud Health Technol Inform 2008;136:597-602. [Medline: 18487796]
33. Preece J, Sharp H, Rogers Y. Interaction Design: Beyond Human-Computer Interaction: John Wiley & Sons; 2015.
34. Scandurra I, Hägglund M, Persson A, Ahlfeldt RM. Disturbing or facilitating?--on the usability of Swedish eHealth systems 2013. Stud Health Technol Inform 2014;205:221-225. [Medline: 25160178]
35. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655. [doi: 10.1136/bmj.a1655] [Medline: 18824488]
36. Scandurra I, Hägglund M, Koch S, Lind M. Usability laboratory test of a novel mobile homecare application with experienced home help service staff. Open Med Inform J 2008;2:117-128. [doi: 10.2174/1874431100802010117] [Medline: 19415140]

37. Scandurra I, Hägglund M, Engström M, Koch S. Heuristic evaluation performed by usability-educated clinicians: education and attitudes. Stud Health Technol Inform 2007;130:205-216. [Medline: 17917194]
38. Scandurra I, Hägglund M, Moström D, Koch S. Heuristic evaluation extended by user analysis: a fast and efficient method to identify potential usability problems in health information systems. J Inf Technol Healthc 2006;4(5):317-325 [FREE Full text]
39. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [doi: 10.1016/s1386-5056(99)00011-8] [Medline: 10530825]
40. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11(1):33. [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
41. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 2010 Nov 16;341:c5814. [doi: 10.1136/bmj.c5814] [Medline: 21081595]
42. Damschroder LJ, Reardon CM, Widerquist MAO, Lowery J. The updated Consolidated Framework for Implementation Research based on user feedback. Implement Sci 2022 Oct 29;17(1):75. [doi: 10.1186/s13012-022-01245-0] [Medline: 36309746]
43. Scrapping the millennium: introduction of a health record in Sweden fails. Heise Online. 2025. URL: https://www.heise.de/en/news/Scrapping-the-millennium-introduction-of-a-health-record-in-Sweden-fails-10323142.html [accessed 2025-08-05]
44. Heggelund MO, Hussain SS, Farshchian BA. Information overload – a case study of using an integrated electronic health record system in the emergency room. SSRN. Preprint posted online on Jun 21, 2024. [doi: 10.2139/ssrn.4866896]
45. Hertzum M, Ellingsen G, Cajander Å. Implementing large-scale electronic health records: experiences from implementations of Epic in Denmark and Finland. Int J Med Inform 2022 Nov;167:104868. [doi: 10.1016/j.ijmedinf.2022.104868] [Medline: 36194994]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
EHR: electronic health record
HCP: health care professionals
i-PARIHS: integrated-Promoting Action on Research Implementation in Health Services
MHCO: municipality home care organization
mHealth: mobile health
PARIHS: Promoting Action on Research Implementation in Health Services
PCC: primary care center

Edited by L Buis; submitted 03.Dec.2024; peer-reviewed by W Odendaal, Y Chu; revised version received 23.Nov.2025; accepted 24.Nov.2025; published 23.Jan.2026.

Please cite as:
Jäderlund Hagstedt L, Hvitfeldt H, Hägglund M
Failed Implementation of Mobile Access to Electronic Health Records in Home Care: Qualitative Study in Sweden
JMIR Mhealth Uhealth 2026;14:e69590
URL: https://mhealth.jmir.org/2026/1/e69590
doi:10.2196/69590

Original Paper

Evaluating the Feasibility of an Electronic Patient-Reported Outcomes Platform Integrating Electronic Health Records and a Mobile Messaging App in Breast Cancer Radiotherapy: Retrospective Cross-Sectional Study

Jong Yun Baek[1*], MD; Haeyoung Kim[1*], MD, PhD; Won Kyung Cho[1], MD; Nalee Kim[1], MD, PhD; Tae Hoon Lee[1], MD, PhD; Won Chul Cha[2,3], MD, PhD

[1]Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea
[2]Department of Digital Health, Samsung Advanced Institute for Health Sciences & Technology, Sungkyunkwan University, Seoul, Republic of Korea
[3]Data Innovation Center, Samsung Medical Center, Seoul, Republic of Korea, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Haeyoung Kim, MD, PhD
Department of Radiation Oncology, Samsung Medical Center, Sungkyunkwan University School of Medicine
81 Irwon-ro, Gangnam-gu, Seoul
Seoul, 06351
Republic of Korea
Phone: 82 2 3410 2612
Fax: 82 2 3410 2619
Email: haeyoung0131.kim@samsung.com

Abstract

Background: Integrating electronic patient-reported outcomes (ePROs) into electronic health records (EHRs) can enhance the quality of patient care. However, collecting longitudinal ePRO data throughout treatment and posttreatment surveillance remains challenging in patients with breast cancer. To address this, we implemented an automated system that enables ePRO acquisition and seamless integration into the EHR. The system delivers questionnaire weblinks via a mobile messaging app, allowing patients to complete ePROs before clinic visits, with responses automatically transferred to the EHR.

Objective: This study aimed to assess patient response rates to the ePRO system and identify key factors influencing the response rate among patients with breast cancer who received radiotherapy and postradiotherapy follow-up.

Methods: We conducted a retrospective analysis of prospectively collected ePRO data by using the BREAST-Q questionnaire, a validated patient-reported outcome measure for breast surgery, from patients who received adjuvant radiotherapy at our institution between May 2023 and April 2024. At a preradiotherapy or postradiotherapy visit, each patient was asked to complete the questionnaire via a weblink sent to their mobile messaging app, KakaoTalk. The questionnaire was dispatched from minutes to several days before each visit. The response rate was calculated as the percentage of patients whose responses were successfully recorded in the EHR among those who were requested to respond. A complete response (CR) was defined as completion of all required questionnaire items. CR rates were analyzed according to clinical factors using univariate and multivariate logistic regression.

Results: Data from 1488 patients were analyzed, encompassing 2431 encounters (median 1, IQR 1-2 per patient). The median age of the patients was 51 (range 23-83) years, with 65.1% (n=968) patients aged 40 to 59 years. Comorbidities were present in 15% (223/1488) of the patients. The CR rate for the first, second, and third ePRO encounters was 89.9% (1338/1488), 98.3% (735/748), and 97.3% (180/185), respectively. Among first-time respondents, younger patients had a significantly higher CR rate (patients aged <60 years: 100/1104, 90.9%; patients aged ≥60 years: 334/384, 87%; *P*=.03). The timing of the questionnaire dispatch also affected the CR rate (*P*<.001). The CR rate was the highest when questionnaires were sent more than 1 hour before the visit (547/583, 93.3%) or in the afternoon of the previous day (505/545, 92.7%) and the lowest when sent 2 or more days before (100/130, 76.9%) or within 1 hour before the appointment (92/112, 81.7%). Both age (*P*=.006) and timing (*P*<.001) remained significant in the multivariate analysis.

Conclusions: This study demonstrates the feasibility of integrating ePRO into EHR through a mobile messaging app–based system, with high patient adherence. Response rates were significantly influenced by patient age and the timing of questionnaire dispatch. These findings provide practical insight for optimizing ePRO implementation in routine oncology care.

(JMIR Mhealth Uhealth 2026;14:e67514) doi:10.2196/67514

KEYWORDS

patient-reported outcomes; mobile app; mHealth; mobile messaging app; electronic health records; eHealth; telemedicine; radiotherapy; breast neoplasms

Introduction

Background

Breast radiotherapy is an essential element in the management of breast cancer, as it enables breast conservation by eliminating microscopic tumor foci following tumor resection and prevents locoregional recurrence, leading to improved survival [1,2]. With advances in multimodality treatment, survival outcomes for patients with breast cancer have substantially improved over the past decades [3,4]. As survival extends, radiotherapy regimens have evolved to minimize treatment-related toxicity while maintaining tumor control [5]. Therefore, capturing toxicity profiles and patient satisfaction is critical for optimizing radiotherapeutic approaches. To achieve this, the incorporation of patient-reported outcomes (PROs) has become increasingly important for individualized counseling and shared decision-making [6].

Despite growing consensus on the value of electronic PROs (ePROs), their routine adoption in oncology remains limited. Collecting longitudinal ePRO data across active treatment and long-term follow-up is particularly challenging, as survivors of breast cancer often require monitoring for more than a decade after completion of primary treatment [7,8]. Previous studies have highlighted several barriers to sustained ePRO use, including workflow burden on clinicians, lack of integration with electronic health records (EHRs), and patient fatigue over repeated reporting [9-12]. Most existing ePRO systems operate as stand-alone or web portal–based platforms, requiring separate log-ins or additional applications, which hinders seamless use during clinical visits [9-11]. As a result, adherence to ePRO completion varies widely, ranging from 27% to 95% across populations with cancer [13-17], and evidence on how to maintain high adherence in daily oncology practice is still lacking.

To address these gaps, we developed and implemented an automated EHR-integrated ePRO system that delivers questionnaires via KakaoTalk (Kakao Corp), the most widely used mobile messaging app in South Korea [18]. This study evaluated the feasibility of this platform by assessing response rates and identifying clinical and contextual factors associated with ePRO adherence among patients receiving postoperative adjuvant breast radiotherapy and follow-up. For an assessment of the feasibility of the ePRO system, the response rate was considered the primary indicator, as it reflects both patient adherence and the sustainability of the platform in routine clinical practice. Consistent with previous ePRO implementation studies [9,10,14,15,17], the response rate has been widely used as a practical measure of feasibility, representing patient engagement, system usability, and operational sustainability within clinical workflows. A consistently high response rate across visits would indicate that the system can be feasibly integrated into long-term follow-up, whereas lower or variable rates would highlight barriers that require further optimization. In addition, factors influencing feasibility were examined by analyzing both clinical and contextual variables that might affect adherence. Clinical variables, such as age, type of surgery, histology, and comorbidities, were considered because they may influence treatment-related experiences and patients' ability to engage with the ePRO systems. Contextual factors, including previous exposure to other ePRO systems and the timing of questionnaire requests, were analyzed because they are expected to directly influence adherence and response behavior.

Study Objective

This study aimed to evaluate the feasibility of integrating an ePRO system directly into the EHR through a mobile messaging app by assessing patient adherence and identifying clinical and contextual factors associated with ePRO completion among patients with breast cancer undergoing postoperative adjuvant radiotherapy and follow-up.

Methods

Participants and Study Design

This study was conducted at the Samsung Medical Center, a large tertiary referral hospital in South Korea, with a comprehensive cancer center located in Seoul. Located in an urban area, the institution serves as a nationwide referral center, providing care to a high volume of patients with cancer from both urban and rural regions. As of 2021, approximately 31,000 patients—representing about 11% of all patients with cancer in South Korea—were treated at the Samsung Medical Center [19]. Between May 2023 and April 2024, ePRO questionnaires were distributed to patients visiting the Department of Radiation Oncology for postoperative adjuvant radiation treatment for breast cancer. These visits included both preradiotherapy evaluations and postradiotherapy follow-ups for surveillance. A weblink connected to the ePRO questionnaires was sent to each individual's KakaoTalk mobile messaging app before their appointment with the attending physician, and patients were asked to respond to the questionnaires through the app. KakaoTalk is a free mobile messaging app used by more than 90% of the Korean population across all age groups and has been widely incorporated into both personal and institutional services. Its functions extend beyond instant messaging to include secure log-in verification, payment systems, and government or health care notifications [20,21]. Because of its

nationwide penetration and user familiarity, the app enables seamless integration of digital health tools, such as ePRO questionnaires, without requiring additional software installation or user registration. The dispatch of the questionnaires was managed by nurses or physicians through a system deployed in the EHR before each patient's visit. The timing of sending the questionnaire varied depending on the sender's preference, ranging from days to minutes before the visit, with no predefined time points. On the day of the visit, an outpatient receptionist or nurse checked whether the questionnaires had been completed. If the questionnaires were incomplete, the receptionist or nurse briefly asked the patient to complete them before the physician's session.

During the study period, the adaptation of the ePRO platform into daily clinical practice varied among the radiation oncologists in our department. Some began using the platform at the beginning of the study period, while others adopted it later. Once a radiation oncologist initiated use of the system, all eligible patients assigned to them were requested to complete the ePRO questionnaires, regardless of individual characteristics. As a result, not all patients with breast cancer visiting our department were uniformly invited to complete ePRO questionnaires. Instead, the number of patients with breast cancer undergoing radiotherapy who were requested to submit ePROs increased toward the latter part of the study period.

After an outpatient visit for preradiotherapy evaluation, patients received radiation treatment according to our institutional protocol. Radiotherapy was administered once daily for 5 consecutive days, with 3 to 19 fractionations over a period of 1 to 4 weeks. Fractionation schedules were determined based on tumor stage, surgery types, the inclusion of regional nodal irradiation, and other risk factors. After completing the treatment, patients were followed up for 2 to 3 weeks after treatment and subsequently every 6 months. ePRO data were collected at the preradiotherapy visit, the immediate postradiotherapy visit at 2 to 3 weeks after treatment, and every 6 months thereafter. As each patient was required to respond to a questionnaire at each hospital visit, 1 or more questionnaires were requested to be completed by each patient during the period of this study.

A System for ePRO Collection and Storage in the EHR

A system for ePRO acquisition and integration with the EHR was developed as an in-house model at our institution. The system links our hospital's EHR with the individual's mobile messaging app for collecting and storing ePRO data. Data entry is performed through the messaging app on the patient's mobile phone, while the data presentation and storage are conducted in our hospital's EHR. A section dedicated to ePRO is integrated into the EHR, allowing physicians and other medical staff to send new ePRO questionnaires on request and view each patient's responses at any time (Figure S1 in Multimedia Appendix 1). When medical staff select ePRO questionnaires and dispatch them through the EHR, a weblink for the ePRO questionnaires is sent to the patient's messaging app. The patient can open the link by entering their date of birth and submit their response to the questionnaire, which is in the form of checkboxes (Figure S2 in Multimedia Appendix 1). The patient's responses are automatically transferred to the ePRO section of the EHR and stored in the hospital's data warehouse. This process of entering, transferring, and storing ePRO data occurs simultaneously in real time, enabling physicians to view the content and time stamp of the data in the EHR.

For the PRO instrument used in this study, we used the Korean version of BREAST-Q (version 2.0) postoperative scale, including modules for breast-conserving therapy, mastectomy, and reconstruction. Among the domains of these modules, the following were used for our patients: satisfaction with breasts, satisfaction with implants, physical well-being of the chest or upper body, and adverse effects of radiation [22,23]. Patients who visited for preradiotherapy evaluation were asked to complete a questionnaire without the domain of adverse effects of radiation, while those attending for postradiotherapy surveillance received a questionnaire that included the domain.

Assessment of the Response Rate and Influencing Factors

Response rate was calculated as the percentage of patients whose responses were successfully recorded in the EHR among those who were requested to respond to the questionnaires. The response rate for each survey encounter was assessed and compared according to the number of encounters, from the first to the third. Because the current analysis was based on surveys conducted over 1 year, most of our patients encountered the questionnaires 1 to 3 times according to the scheduled follow-up interval. We classified response status into 3 categories: complete response (CR; all questions answered), partial response (PR; at least 1 question answered but not all), and no response (NR). In addition, when analyzing significant factors influencing CR, we divided our patients into 2 groups: complete responders and noncomplete responders, with partial responders and nonresponders merged into the noncomplete responder group.

To determine the significant factors influencing CR, we compared the CR rate according to various factors, focusing only on patients who encountered the survey for the first time. Specifically, to assess the impact of questionnaire request timing before a visit appointment, the timing was categorized into 5 groups: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before). In addition, patients who had previous experiences responding to questionnaires requested from other departments in our hospital were categorized as previous other ePRO (+), while those without the experiences were categorized as previous other ePRO (–). Finally, patients with any of the following diseases were categorized as having comorbidity: diabetes, cardiovascular disease, chronic pulmonary disease, hepatic disease, renal disease, or other cancers.

Statistical Analysis

For the univariate analysis, Fisher exact test was used to compare the CR according to clinical and contextual variables, including age, type of surgery, histology, comorbidities, previous

ePRO experience, and the timing of questionnaire requests. Age was categorized using a 60-year cutoff for univariate analysis and treated as a continuous variable in multivariate models. Patients with missing data for a given variable were excluded from the study. Factors with $P<.10$ in univariate analyses were included in the multivariate logistic regression model to adjust for potential confounding effects. All P values were 2-sided, with $P<.05$ considered statistically significant. All analyses were performed using SPSS software (version 27; IBM Corp).

Ethical Considerations

This study was approved by the institutional review board of the Samsung Medical Center (institutional review board approval number 2024-07-147-001). The analysis was conducted retrospectively using deidentified ePRO data collected during routine clinical care, and the need for informed consent was waived by the institutional review board. All patient information was deidentified before analysis to ensure privacy and confidentiality. No personal identifiers were included in the downloaded dataset. No compensation was provided to patients for participation, as the data were collected as part of standard clinical procedures without additional burden or intervention.

Results

Patients' Characteristics

A total of 2334 patients with breast cancer attended our department during the study period. Of the 2334 patients, 1491 (63.9%) with nonmetastatic breast cancer were invited to complete the ePRO survey. This difference reflects the gradual adoption of the ePRO system among physicians, as the platform was implemented in stages and not all attending radiation oncologists had begun using it at the start of the study period. A total of 3 (0.2%) patients were excluded from the analysis due to missing time records of their questionnaire responses, resulting in 1488 (99.8%) patients being included in this study. Among them, 740 (49.7%) encountered 1 survey, 563 (37.8%) encountered 2 surveys, 175 (11.8%) encountered 3 surveys, and 10 (0.7%) encountered 4 surveys, resulting in 2431 survey encounters.

The characteristics of the 1488 patients are summarized in Table 1. The median age was 51 years, with most of the patients (n=968, 65.1%) aged between 40 and 59 years, and 41 (2.8%) patients were aged 75 years or older. Most patients underwent breast-conserving surgery (n=1186, 79.7%) and had invasive carcinoma (n=1342, 90.2%). Comorbidities were found in 223 (14.9%) patients, including diabetes (n=98, 6.6%); cardiovascular disease (n=67, 4.5%); chronic disease of the liver, lung, or kidney (n=63, 4.2%); and other cancers (n=26, 1.7%).

Table 1. Patients' characteristics (N=1488).

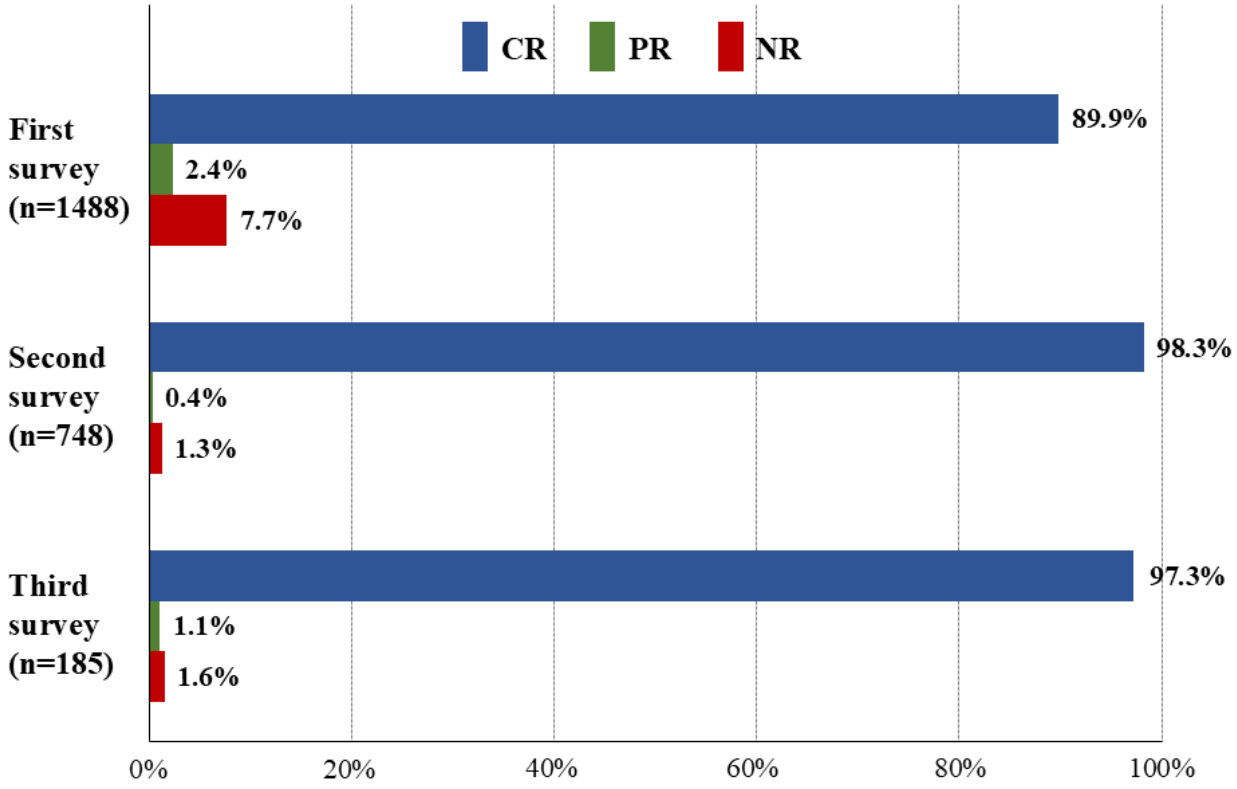
Characteristics	Patients
Age (y), median (range)	51 (24-85)
Age (y), n (%)	
≥20 to <40	136 (9.1)
≥40 to <60	968 (65.1)
≥60	384 (25.8)
Sex, n (%)	
Female	1487 (99.9)
Male	1 (0.1)
Type of surgery, n (%)	
Breast-conserving surgery	1186 (79.7)
Mastectomy without reconstruction	129 (8.7)
Mastectomy with reconstruction	173 (11.6)
Histology, n (%)	
Ductal carcinoma in situ	146 (9.8)
Invasive carcinoma	1342 (90.2)
Type of visit, n (%)	
Preradiotherapy evaluation visit	946 (63.6)
Postradiotherapy follow-up visit	542 (36.4)
Comorbidity, n (%)	
Yes	223 (15)
No	1265 (85)
Experience of other electronic patient-reported outcomes, n (%)	
Yes	392 (26.3)
No	1096 (73.7)
Timing of the questionnaire requests[a], n (%)	
≤1 h on the day	115 (7.7)
>1 h on the day	586 (39.4)
PM the day before	545 (36.6)
AM the day before	112 (7.5)
≥2 d before	130 (8.7)

[a]The timing of the questionnaire requests before a visit appointment was categorized as follows: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before).

Response Rate and Influencing Factors

Of the 1488 patients who encountered the questionnaires for the first time, 1338 (89.9%) exhibited CR, 35 (2.4%) submitted PR, and 115 (7.7%) did not respond to the questionnaire. The response status and rates according to the number of survey encounters are shown in Figure 1.

Figure 1. Rates of complete response (CR), partial response (PR), and no response (NR) for electronic patient-reported outcomes surveys across first, second, and third requests.



In the univariate analysis of factors influencing CR, the following were statistically significant: patient's age, type of visit, and the timing of the questionnaire requests before a visit appointment (Table 2). The CR rate was 90.9% (1004/1104) in patients younger than 60 years and 87% (334/384) in those aged 60 years or older (P=.03), indicating that older age was associated with lower completion. A significantly lower CR rate was also observed in patients receiving the questionnaire at the preradiotherapy visit (837/946, 88.5%) compared to the postradiotherapy visits (501/542, 92.4%; P=.02).

Table 2. Univariate analysis of factors affecting the complete response to the electronic patient-reported outcome questionnaire (N=1488).

Characteristics	CR[a], n (%)	Non-CR, n (%)	*P* value
Age (y)			.03
<60	1004 (90.9)	100 (9.1)	
≥60	334 (87)	50 (13)	
Type of surgery			.10
Breast-conserving surgery	1057 (89.1)	129 (10.9)	
Mastectomy without reconstruction	122 (94.6)	7 (5.4)	
Mastectomy with reconstruction	159 (91.9)	14 (8.1)	
Histology			.77
Ductal carcinoma in situ	133 (91.1)	13 (8.9)	
Invasive carcinoma	1205 (89.8)	137 (10.2)	
Type of visit			.02
Preradiotherapy evaluation visit	837 (88.5)	109 (11.5)	
Postradiotherapy follow-up visit	501 (92.4)	41 (7.6)	
Comorbidity			.55
Yes	198 (88.8)	25 (11.2)	
No	1140 (90.1)	125 (9.9)	
Experience of other electronic patient-reported outcome			.85
Yes	354 (90.3)	38 (9.7)	
No	984 (89.8)	112 (10.2)	
			<.001
≤1 h on the day	94 (81.7)	21 (18.3)	
>1 h on the day	547 (93.3)	39 (6.7)	
PM the day before	505 (92.7)	40 (7.3)	
AM the day before	92 (82.1)	20 (17.9)	
≥2 d before	100 (76.9)	30 (23.1)	

[a]CR: complete response.

[b]The timing of the questionnaire requests before a visit appointment was categorized as follows: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before).

Notably, the timing of the questionnaire requests had a strong influence on CR. Patients who received the questionnaire more than 1 hour before the appointment or in the afternoon of the previous day showed the highest CR rates at 93.3% (547/583) and 92.7% (505/545), respectively. In contrast, the lowest rates were seen in those who received the questionnaire 2 or more days before (100/130, 76.9%) and within 1 hour of the appointment (94/115, 81.7%; *P*<.001).

In the multivariate analysis, age and the timing of the questionnaire requests remained significant influencing factors for CR. Old age (odds ratio 0.98, 95% CI 0.96-0.99; *P*=.006) and questionnaire request timing of ≤1 hour on the day of the visit, AM the day before, or ≥2 days before the appointment (odds ratio 0.32, 95% CI 0.22-0.45; *P*<.001) were significantly associated with a lower CR rate (Table 3).

Table 3. Multivariate analysis of factors affecting the complete response to the electronic patient-reported outcome questionnaire.

Characteristics	Odds ratio (95% CI)	*P* value
Age (y): continuous	0.98 (0.96-0.99)	.006
Type of surgery: breast-conserving surgery vs mastectomy or reconstruction	0.64 (0.39-1.05)	.08
Type of visit: preradiotherapy evaluation visit vs postradiotherapy follow-up visit	1.33 (0.90-1.97)	.15
Timing of the questionnaire requests[a]: ≤1 h on the day or AM the day before or ≥2 d before vs >1 h on the day or PM the day before	0.32 (0.22-0.45)	<.001

[a]The timing of ePRO requests before a visit appointment was categorized as follows: within 1 hour of the appointment time (≤1 hour on the day), more than 1 hour before the appointment on the visit day (>1 hour on the day), in the afternoon of the day before the appointment (PM the day before), in the morning of the day before the appointment (AM the day before), and 2 or more days before the appointment day (≥2 days before).

Discussion

Principal Findings

Patients using our ePRO system, which links a commercial mobile messaging app with our hospital's EHR, showed an 89.9% (1338/1488) CR rate for the BREAST-Q questionnaires. Responses of patients who visited our radiation oncology department for breast cancer management were successfully recorded. The response rate to the questionnaires increased as the number of survey encounters increased. Age of 60 years or older was associated with a lower rate of CR; however, 87% (334/384) of participants of this age group provided appropriate responses to the questionnaires delivered through their mobile messaging app, even if it was their first time encountering the questionnaires using the app. Notably, the timing of the questionnaire requests significantly influenced the CR rate, with a higher CR rate of more than 92% observed when the questionnaires were requested more than 1 hour before the scheduled visit or in the afternoon of the day before the appointment. Given these findings, our ePRO system shows potential as a feasible platform for ePRO collection and integration with the hospital's EHR. In addition, the factors identified as significantly affecting the CR rate of the questionnaires can be used to guide improvements in responses to ePRO questionnaires in daily practice.

Comparison With Prior Work

The significance of PRO measurement in oncology care has been increasingly emphasized in recent years [24]. In this regard, the European Society for Medical Oncology released a clinical practice guideline concerning the use of PRO in the continuum of cancer care, emphasizing the essential role of symptom monitoring via PRO measurements [25]. ePRO offers several advantages, such as greater patient preference and acceptability, lower human resource costs, and higher data quality [9,11]. Various forms of ePRO collection platforms, including web-based and app-based systems, have been developed and used [13,14,26]. In addition to the data collection system, integrating the data into the EHR is essential to facilitate the incorporation of ePRO into clinical practice [9-12].

In this analysis, we found favorable patient adherence to our ePRO system, with more than 89.9% (1338/1488) CR rate for the ePRO questionnaires, even at the first encounter. This promising result is attributed to adopting a system that uses the KakaoTalk messaging app, which was familiar to our patients [18]. Patients were able to access the questionnaires using the existing app on their smartphone, without the need to install an additional app for ePRO. Because the message with the questionnaire link was sent under the hospital's name, our patients likely accepted it confidently, without concerns about cybercrimes. Moreover, given that more than 94% of Koreans own a smartphone, ePRO questionnaires delivered via the mobile app could effectively encourage responses from our patient population [27]. According to previous studies, ePRO adherence rate among patients with cancer has been reported to range between 27% and 95% [13-17]. In a study conducted in the United States, ePRO adherence rates ranged from 27% to 70%, following administration either on-site via tablet or remotely via a patient portal. There were significant differences in response rates depending on patient age and race, with older patients aged 65 years and older and non-White individuals being negatively associated with adherence to ePRO [14]. Meanwhile, a Japanese study reported a 95% ePRO adherence rate via a mobile messaging app, LINE, from 40 participants, which was similar to our patients' adherence rate [16]. Considering that LINE is used by more than 70% of the Japanese population, the familiarity with the ePRO acquisition tool likely contributed to the high ePRO adherence in their study [28]. Taken together, these findings suggest that selecting appropriate tools for ePRO administration based on respondents' demographic or cultural characteristics is essential for achieving favorable ePRO acceptance.

Older age has been reported to be significantly associated with lower adherence to ePRO [13,14,17]. In a prospective study conducted among French patients aged 75 years and older, 26% of the participants accepted ePRO, which was conducted remotely using a web-based app [13]. More than 52% of the participants did not respond to the ePRO due to technological issues, such as a lack of internet access or discomfort with using the internet [13]. In addition, a study performed in the United States showed that patients aged 65 years or older exhibited a 6% decrease in adherence to ePRO, which was a significant difference compared to younger patients [14]. Similarly, in our study, patients aged 60 years or older showed a significantly lower CR rate for ePRO than those younger than 60 years. However, considering that 87% (334/384) of the older participants completed the questionnaires, the ePRO acceptance using our system appears favorable even among older patients. In our study, ePRO was requested from all patients attending

our department, without selection based on their smartphone possession or daily internet use. Furthermore, given that more than 90% of Koreans aged 60 years or older use smartphones [29,30] and most are reported to be familiar with KakaoTalk [31], the ePRO acceptance among our older participants likely reflects the real-world feasibility of implementing ePRO in clinical practice for older Korean patients, particularly when it is delivered through the familiar messaging app. In the meantime, we also found that 13% of patients in this older age group did not properly respond to ePRO, suggesting that there is room for improvement in enhancing ePRO adherence among older patients. It is uncertain why these patients did not respond to the ePRO measurements, as we did not assess the reasons for nonresponse to the questionnaires. However, referencing previous studies, various factors have been identified that affect ePRO acceptance in older patients, including frailty level, socioeconomic status, technological barriers, and the modes of ePRO administration [13,30,32]. Future studies are needed to determine the causes of nonadherence to ePRO and to provide the most appropriate ePRO collection modalities based on the individual characteristics of older Korean patients undergoing cancer treatments.

Optimizing ePRO Response Through Timing

Interestingly, we found that the timing of ePRO requests was significantly related to the patient's CR rate. The most appropriate time for requesting ePRO questionnaires was either more than 1 hour before the appointment or in the afternoon on the day before the scheduled visit. This finding suggests that patients may feel more comfortable and have sufficient time to review and respond to ePRO requests when they are delivered within this timeframe. Delivering ePRO questionnaires more than 2 days before a scheduled visit may have hindered appropriate responses, possibly due to difficulties in locating our ePRO request message among other personal messages. As our patients' ePROs were collected remotely using a mobile app and their completeness was rechecked on-site in our clinic, it is likely that our ePRO acceptance rate is higher than that reported in settings relying solely on remote ePROs collection. This may be indirectly supported by the CR rate of 76.9% (100/130) among our patients who received ePRO questionnaires more than 2 days before an appointment. However, from another perspective, the combination of remote ePRO collection via a messaging app, requested within a specific timeframe, and on-site feedback appears to be an effective strategy for maximizing ePRO acceptance, as indicated by a CR rate of more than 93.8% (547/583) among our patients.

Strengths and Limitations

A key strength of this study is the provision of real-world data on ePRO adherence and the identification of significant factors influencing adherence among 1488 patients with nonmetastatic breast cancer following implementation of an in-house ePRO platform. Because our ePRO platform uses a messaging app that is largely familiar to Koreans, we observed a favorable acceptance for the ePRO. However, we acknowledge the limitations of this study. Our data were retrospectively derived from a single institution over a 1-year data collection period. Therefore, the data may be insufficient to capture long-term ePRO adherence among general patients with breast cancer. In addition, reasons for PR or NR to ePRO questionnaires were not available, as this analysis was conducted retrospectively. Because the causes of incomplete ePRO are important for identifying areas of improvement in enhancing ePRO adherence, further assessments are necessary among those submitting PR or NR to ePRO questionnaires in future studies. Furthermore, because our ePRO platform relies on a specific mobile messaging app—KakaoTalk, which is predominantly used in South Korea—its general acceptance in other countries remains uncertain. Therefore, the generalizability of our findings to other populations with different digital habits may be limited. In addition, technological barriers or disparities in mobile access may have also influenced patient participation and response accuracy. Moreover, as all our ePRO data were self-reported, reporting bias may have been introduced. Future research involving institutions that use different platforms or serve more diverse populations is needed to explore the generalizability of our findings.

Conclusions

The collection of ePRO data and its integration into EHR was successful with our ePRO platform, achieving an overall CR rate of 89.9% (1338/1488). Patient age and the specific timeframe for ePRO requests were significant factors influencing complete ePRO acceptance. Patients aged 60 years or older showed significantly lower ePRO adherence. In addition, a specific timeframe, including more than 1 hour before a clinical visit or in the afternoon on the day before the appointment, was associated with a significantly higher CR rate for ePRO. These factors are expected to improve ePRO acceptance among patients with nonmetastatic breast cancer.

Funding

This study was supported by grants from the National Research Foundation of Korea funded by the Ministry of Education (RS-2023-00252283) and from the Samsung Medical Center (SMO1240501).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HK, WKC, and NK planned the study. HK, WKC, NK, and THL conducted the electronic patient-reported outcomes survey. JYB carried out data collection and analysis. JYB and HK drafted the manuscript. WCC designed the electronic patient-reported outcomes system. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots illustrating the workflow of the electronic health record–integrated electronic patient-reported outcome system and the patient questionnaire interface.
[DOCX File , 522 KB - mhealth_v14i1e67514_app1.docx]

References

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Darby S, McGale P, Correa C, Taylor C, Arriagada R, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: meta-analysis of individual patient data for 10,801 women in 17 randomised trials. Lancet 2011 Nov 12;378(9804):1707-1716 [FREE Full text] [doi: 10.1016/S0140-6736(11)61629-2] [Medline: 22019144]
2. Early Breast Cancer Trialists' Collaborative Group (EBCTCG). Radiotherapy to regional nodes in early breast cancer: an individual patient data meta-analysis of 14 324 women in 16 trials. Lancet 2023 Nov 25;402(10416):1991-2003 [FREE Full text] [doi: 10.1016/S0140-6736(23)01082-6] [Medline: 37931633]
3. Wojtyla C, Bertuccio P, Wojtyla A, La Vecchia C. European trends in breast cancer mortality, 1980-2017 and predictions to 2025. Eur J Cancer 2021 Jul;152:4-17 [FREE Full text] [doi: 10.1016/j.ejca.2021.04.026] [Medline: 34062485]
4. Giaquinto AN, Sung H, Miller KD, Kramer JL, Newman LA, Minihan A, et al. Breast cancer statistics, 2022. CA Cancer J Clin 2022 Nov 03;72(6):524-541 [FREE Full text] [doi: 10.3322/caac.21754] [Medline: 36190501]
5. Shah C, Al-Hilli Z, Vicini F. Advances in breast cancer radiotherapy: implications for current and future practice. JCO Oncol Pract 2021 Dec;17(12):697-706. [doi: 10.1200/OP.21.00635] [Medline: 34652952]
6. Alcorn SR, Corbin KS, Shumway DA. Integrating the patient's voice in toxicity reporting and treatment decisions for breast radiotherapy. Semin Radiat Oncol 2022 Jul;32(3):207-220. [doi: 10.1016/j.semradonc.2022.01.010] [Medline: 35688519]
7. Pedersen RN, Esen BO, Mellemkjær L, Christiansen P, Ejlertsen B, Lash TL, et al. The incidence of breast cancer recurrence 10-32 years after primary diagnosis. J Natl Cancer Inst 2022 Mar 08;114(3):391-399 [FREE Full text] [doi: 10.1093/jnci/djab202] [Medline: 34747484]
8. De Rose F, Meduri B, De Santis MC, Ferro A, Marino L, Colciago RR, et al. Rethinking breast cancer follow-up based on individual risk and recurrence management. Cancer Treat Rev 2022 Sep;109:102434 [FREE Full text] [doi: 10.1016/j.ctrv.2022.102434] [Medline: 35933845]
9. Cracchiolo JR, Arafat W, Atreja A, Bruckner L, Emamekhoo H, Heinrichs T, et al. Getting ready for real-world use of electronic patient-reported outcomes (ePROs) for patients with cancer: a National Comprehensive Cancer Network ePRO Workgroup paper. Cancer 2023 Aug 15;129(16):2441-2449 [FREE Full text] [doi: 10.1002/cncr.34844] [Medline: 37224181]
10. Hassett MJ, Cronin C, Tsou TC, Wedge J, Bian J, Dizon DS, et al. eSyM: an electronic health record-integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. JCO Clin Cancer Inform 2022 Jan;6:e2100137 [FREE Full text] [doi: 10.1200/CCI.21.00137] [Medline: 34985914]
11. Meirte J, Hellemans N, Anthonissen M, Denteneer L, Maertens K, Moortgat P, et al. Benefits and disadvantages of electronic patient-reported outcome measures: systematic review. JMIR Perioper Med 2020 Apr 03;3(1):e15588 [FREE Full text] [doi: 10.2196/15588] [Medline: 33393920]
12. Salmani H, Nasiri S, Ahmadi M. The advantages, disadvantages, threats, and opportunities of electronic patient-reported outcome systems in cancer: a systematic review. Digit Health 2024;10:20552076241257146 [FREE Full text] [doi: 10.1177/20552076241257146] [Medline: 38812853]
13. Cancel M, Sauger C, Biogeau J, Dardaine-Giraud V, Lecomte T, Solub D, et al. FASTOCH: feasibility of electronic patient-reported outcomes in older patients with cancer-a multicenter prospective study. J Clin Oncol 2024 Aug 01;42(22):2713-2722. [doi: 10.1200/JCO.23.02150] [Medline: 38709983]
14. Takvorian SU, Anderson RT, Gabriel PE, Poznyak D, Lee S, Simon S, et al. Real-world adherence to patient-reported outcome monitoring as a cancer care quality metric. JCO Oncol Pract 2022 Sep;18(9):e1454-e1465. [doi: 10.1200/OP.21.00855] [Medline: 35675586]
15. Takala L, Kuusinen TE, Skyttä T, Kellokumpu-Lehtinen P, Bärlund M. Electronic patient-reported outcomes during breast cancer adjuvant radiotherapy. Clin Breast Cancer 2021 Jun;21(3):e252-e270 [FREE Full text] [doi: 10.1016/j.clbc.2020.10.004] [Medline: 33229222]

16. Hayashida T, Nagayama A, Seki T, Takahashi M, Matsumoto A, Kubota A, et al. Feasibility study on collecting patient-reported outcomes from breast cancer patients using the LINE-ePRO system. Cancer Sci 2022 May;113(5):1722-1730 [FREE Full text] [doi: 10.1111/cas.15329] [Medline: 35279907]
17. El Shafie RA, Weber D, Bougatf N, Sprave T, Oetzel D, Huber PE, et al. Supportive care in radiotherapy based on a mobile app: prospective multicenter survey. JMIR Mhealth Uhealth 2018 Aug 30;6(8):e10916 [FREE Full text] [doi: 10.2196/10916] [Medline: 30166275]
18. Digital 2024: South Korea. Kepios. URL: https://datareportal.com/reports/digital-2024-south-korea [accessed 2025-05-29]
19. Cancer center outcomes and introduction. Samsung Medical Center Cancer Hospital. URL: http://www.samsunghospital.com/home/cancer/intro.do?view=OUTCOMES [accessed 2025-05-29]
20. KakaoTalk overview. Kakao Corp. URL: https://www.kakaocorp.com/service/KakaoTalk?lang=en [accessed 2026-01-19]
21. Kim B. South Korea's Megacorp and super app: Kakao's paths to market dominance. Media Cult Soc 2024 Nov 16;47(4):641-659. [doi: 10.1177/01634437241294207]
22. Seth I, Seth N, Bulloch G, Rozen WM, Hunter-Smith DJ. Systematic review of Breast-Q: a tool to evaluate post-mastectomy breast reconstruction. Breast Cancer (Dove Med Press) 2021;13:711-724 [FREE Full text] [doi: 10.2147/BCTT.S256393] [Medline: 34938118]
23. Liu LQ, Branford OA, Mehigan S. BREAST-Q measurement of the patient perspective in oncoplastic breast surgery: a systematic review. Plast Reconstr Surg Glob Open 2018 Aug;6(8):e1904 [FREE Full text] [doi: 10.1097/GOX.0000000000001904] [Medline: 30254830]
24. Howell D, Molloy S, Wilkinson K, Green E, Orchard K, Wang K, et al. Patient-reported outcomes in routine cancer clinical practice: a scoping review of use, impact on health outcomes, and implementation factors. Ann Oncol 2015 Sep;26(9):1846-1858 [FREE Full text] [doi: 10.1093/annonc/mdv181] [Medline: 25888610]
25. Di Maio M, Basch E, Denis F, Fallowfield LJ, Ganz PA, Howell D, ESMO Guidelines Committee. The role of patient-reported outcome measures in the continuum of cancer clinical care: ESMO Clinical Practice guideline. Ann Oncol 2022 Sep;33(9):878-892 [FREE Full text] [doi: 10.1016/j.annonc.2022.04.007] [Medline: 35462007]
26. Basch E, Schrag D, Henson S, Jansen J, Ginos B, Stover AM, et al. Effect of electronic symptom monitoring on patient-reported outcomes among patients with metastatic cancer: a randomized clinical trial. JAMA 2022 Jun 28;327(24):2413-2422 [FREE Full text] [doi: 10.1001/jama.2022.9265] [Medline: 35661856]
27. Yoon JS. Smartphone usage in South Korea - statistics and facts. Statista. URL: https://www.statista.com/topics/5340/smartphone-usage-in-south-korea/#topicOverview [accessed 2025-05-29]
28. Why is LINE the most popular social media app in Japan? Digital Marketing for Asia. URL: https://www.digitalmarketingforasia.com/why-line-is-the-most-popular-social-media-app-in-japan/ [accessed 2025-05-29]
29. Kang H, Baek J, Chu SH, Choi J. Digital literacy among Korean older adults: a scoping review of quantitative studies. Digit Health 2023;9:20552076231197334 [FREE Full text] [doi: 10.1177/20552076231197334] [Medline: 37654708]
30. Lee H, Choi JY, Kim SW, Ko KP, Park YS, Kim KJ, et al. Digital health technology use among older adults: exploring the impact of frailty on utilization, purpose, and satisfaction in Korea. J Korean Med Sci 2024 Jan 08;39(1):e7 [FREE Full text] [doi: 10.3346/jkms.2024.39.e7] [Medline: 38193326]
31. Kim K, Lee B, Park Y, Jung EY, Kim SK, Suh DH, et al. Factors encouraging mobile instant messaging service use in medical education. PeerJ 2019;7:e7275 [FREE Full text] [doi: 10.7717/peerj.7275] [Medline: 31328039]
32. Appleyard SE, Larkin MJ, Stewart EM, Minton O, Gilbert DC. Digital medicine in men with advanced prostate cancer - a feasibility study of electronic patient-reported outcomes in patients on systemic treatment. Clin Oncol (R Coll Radiol) 2021 Dec;33(12):751-760. [doi: 10.1016/j.clon.2021.04.008] [Medline: 33966948]

Abbreviations

CR: complete response
EHR: electronic health record
ePRO: electronic patient-reported outcome
NR: no response
PR: partial response
PRO: patient-reported outcome

Edited by L Buis; submitted 14.Oct.2024; peer-reviewed by ME Arayici, KL Mauco; comments to author 01.Apr.2025; accepted 16.Jan.2026; published 05.Feb.2026.

Please cite as:
Baek JY, Kim H, Cho WK, Kim N, Lee TH, Cha WC
Evaluating the Feasibility of an Electronic Patient-Reported Outcomes Platform Integrating Electronic Health Records and a Mobile Messaging App in Breast Cancer Radiotherapy: Retrospective Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e67514
URL: https://mhealth.jmir.org/2026/1/e67514
doi:10.2196/67514
PMID:

Impact of Mobilization Facilitated by Wearable Device Enhanced Patient Monitoring/Electrophysiology Pod–Based Feedback on Postoperative Complications Following Colorectal Cancer Surgery: Randomized Controlled Trial

Yang Meng[1*], MS; Fengyan Fan[1*], BS; Yumei Ma[2*], MS; Nong Yan[3], MS; Huan Wang[1], MS; Zhen Zhang[1], MS; Yiting Wang[1], MS; Hailong Dong[1], MD, PhD; Huang Nie[1], MD, PhD

[1]Department of Anesthesiology and Perioperative Medicine, Fourth Military Medical University Xijing Hospital, No.127 Changle West Road, Xi'an, China

[2]Department of Anesthesiology, No.905 Hospital of PLA NAVY, Shanghai, China

[3]Mindray Medical International Ltd, Shenzhen, China

[*]these authors contributed equally

Corresponding Author:
Huang Nie, MD, PhD
Department of Anesthesiology and Perioperative Medicine, Fourth Military Medical University Xijing Hospital, No.127 Changle West Road, Xi'an, China

Abstract

Background: Enhanced recovery after surgery (ERAS) guidelines recommend early postoperative mobilization to reduce complications, but adherence is often suboptimal, highlighting the need for effective tools to monitor and encourage movement. The Mindray enhanced patient monitoring (ePM)/electrophysiology (ep) pod, capable of tracking activity, vital signs, sleep, and pain, offers high-precision postoperative monitoring and is well-suited for research on activity feedback.

Objective: The study aims to assess whether wearable device-based (ePM/ep pod) activity feedback could reduce postoperative complications within 30 days of colorectal cancer (CRC) surgery.

Methods: We conducted an open-label, evaluator-blind, randomized controlled trial involving patients aged ≥18 years scheduled for CRC surgery. Patients were randomized to a feedback group or a control group. Both groups were set the same target activity time postoperatively based on ERAS guidelines. The feedback group received real-time visual feedback of movement time daily through the ePM/ep pod device, while the control group did not receive feedback. The primary outcome was the comprehensive complication index (CCI) within postoperative 30 days. Secondary outcomes included daily activity time, pain Numeric Rating Scale scores for rest and movement during the first 3 postoperative days, length of stay, percentage of reaching the scheduled mobilization target, 30-day postoperative mortality rate, and the times of first exhaust and defecation.

Results: Two hundred thirty-nine patients were recruited between February 2023 and September 2023, with 216 randomized (n=108 for each group). There was no significant difference in CCI within 30 postoperative days between the control group (median CCI 0, range 0 - 20.90) and the activity feedback group (median CCI 0, range 0 - 12.20). The estimated mean difference was −0.59 (95% CI −3.56 to 2.38; *P*=.66). Sensitivity analysis excluding patients with low device compliance did not alter these findings. No significant differences between groups were found in daily activity time, length of hospital stay, or pain scores. Post hoc analysis revealed significant negative correlations between 30-day CCI and activity on the second day after operation (*r*=−0.166) and the third day after operation (POD3) (*r*=−0.264; *P*<.05 for both). Linear regression indicated that POD3 activity significantly reduced CCI (β=−.025; 95% CI −0.045 to −0.006*; P*=.01), with peak CCI reduction at 215 minutes of activity.

Conclusions: In the context of ERAS, this study found no evidence that activity stimulation based on feedback from the wearable device (ePM/ep pod) could reduce 30-day postoperative CCI in patients undergoing CRC surgery. However, the ePM/ep pod could accurately record daily activity duration, which may be negatively correlated with CCI on POD3.

Trial Registration: Chinese Clinical Trial Register ChiCTR2300068107; https://www.chictr.org.cn/showproj.html?proj=189756

(JMIR Mhealth Uhealth 2026;14:e70534) doi:10.2196/70534

KEYWORDS

wearable device; postoperative complications; ERAS; colorectal cancer; postoperative mobilization; enhanced recovery after surgery

Introduction

Colorectal cancer (CRC) is the second most common cancer in China and ranks third globally in mortality rate [1,2], which is an urgent health problem. Surgery is the first-line treatment of CRC. Unfortunately, 50% of patients undergoing surgery develop postoperative complications [3], leading to increased medical expenses, longer hospital stays, and delay of postoperative recovery [4]. Previous studies have shown that postoperative complications are an independent risk factor related to poor 5-year survival and the increased overall recurrence rate [5-7].

Enhanced recovery after surgery (ERAS) is committed to reducing surgical trauma and stress, decreasing postoperative morbidity, and accelerating postoperative recovery of patients [8-10]. ERAS guidelines emphasize the importance of early mobilization after surgery [11], which can reduce adverse outcomes such as pulmonary complications, decreased skeletal muscle strength, thromboembolism, and insulin resistance [12]. It was confirmed that ERAS was safe and effective in patients undergoing CRC surgery [13,14], and early mobilization after operation was the key to the success of ERAS [15]. However, more than half of the patients did not mobilize as recommended in the ERAS guidelines [16], which meant a reliable measurement tool was needed to not only accurately monitor the postoperative activity of patients, but also to increase their enthusiasm for mobilization. One of the possibilities is to use a wearable device [17,18].

The development of wearable devices allows objective measurements of, for example, activity steps or activity time, which can be used as a feedback tool to urge patients to achieve their intended mobilization goals [17-19]. A study found it was safe to use an activity tracker with feedback after gynecological operations and may increase patients' physical activity [20]. Several wearable devices (such as Apple Watch, Fitbit, and Garmin) that provide combined automatic and visual feedback have been identified as effective tools for cancer survivors to raise awareness of self-activity, stimulate behavior change, and improve physical activity [21,22]. However, the clinical effect on recovery or promoting movement after operation of the wearable devices has not been sufficiently evaluated in patients with CRC. The enhanced patient monitoring (ePM)/electrophysiology (ep) pod developed by Mindray supports activity time statistics, continuous vital sign monitoring, sleep monitoring, pain scores, and trend recording, has a high accuracy of postoperative activity monitoring [23], and is suitable for related research.

The present study aims to determine whether the activity incentive feedback based on a wearable device (ePM/ep pod) could reduce postoperative complications of patients with CRC within 30 days after their operations. We hypothesized that postoperative complications would be reduced in the activity feedback group compared to the usual postoperative care group.

Methods

Study Design and Patients

It was an open-label, evaluator-blind, randomized controlled study with 2 parallel groups.

Patients aged ≥18 years and scheduled for CRC surgery were included. The exclusion criteria were patients who (1) had American Society of Anesthesiologists (ASA) grade >III, (2) had preoperative immobility or inability to walk unaided affected by neural, muscular, or skeletal diseases, (3) had participated in other clinical studies in the past 3 months, (4) had planned reoperation within 30 days after the initial operation, and (5) had expected to be transferred to the intensive care unit after their operations.

Ethical Considerations

The Ethics Committee of the First Affiliated Hospital of the Air Force Military Medical University approved this randomized controlled trial (approval no. KY20222299-F-1), and the study protocol was registered with the Chinese Clinical Trial Registry (ChiCTR2300068107) on July 2, 2023. All procedures complied with the Declaration of Helsinki and institutional regulations (The Ethics Committee of the First Affiliated Hospital of the Air Force Military Medical University, Medical Ethics Committee Charter) [24]. Written informed consent was obtained from all participants prior to enrollment, which included detailed explanation of the study and their right to withdraw at any time. Upon registration, each patient was assigned a unique study number, and all subsequent data analyses were performed using a deidentified dataset to ensure participant privacy and confidentiality. No financial compensation was provided for participation.

Interventions

All patients in the trial were randomly assigned (1:1) to a feedback group or a control group. Both groups were set the same target activity time postoperatively and thus performed activities for 2 hours on the first day after operation (POD1), 4 hours on the second day after operation (POD2), and 6 hours on the third day after operation (POD3), according to the ERAS guidelines and our ERAS pathway for CRC [10]. All patients were fitted with a wireless wearable device that continuously recorded their daily ambulation durations (ePM 12M monitor placed at the bedside or ep pod monitor placed on their wrist or blood pressure pod monitor placed on their upper arm).

In the feedback group, patients received real-time visual feedback of the movement time every day through the display of the ePM. According to the wards' routine care based on ERAS guidelines, patients were reminded to engage in activities combined with ePM/ep pod feedback to meet predefined daily step goals and also informed about the gap between the active time and the target time.

In the control group, patients' movement time was set to not appear on the display of the ePM. Patients were not given feedback, and they were only encouraged to carry out activities according to the current care standards based on ERAS guidelines.

Prevention and Handling of Adverse Events

If the following incidents occur, activity supervision should be suspended immediately: participants experience pain intensity >4 (from 0 to -10), electrocardiogram abnormalities (signs of arrhythmia or acute ischemia), indoor air oxygen saturation <88%, systolic blood pressure <90 mmHg, symptoms of standing intolerance (ie, dizziness, nausea, and blurred vision), exercise weakness (ie, difficulty standing), etc. In the next scheduled mobilization, if the participant is considered to be in stable physical condition after clinical evaluation, a new mobilization attempt will be conducted. If serious adverse events such as falling out of bed or falling occur, in addition to suspending activity supervision, the competent physician should also be asked to evaluate the patient's injury and terminate the study if necessary. The researcher should immediately fill out the serious adverse event report and report it to the hospital ethics committee.

Randomization and Blinding

Patients were randomized using block random groups (sized blocks of 2 and 4) generated by specialized staff. All random assignments were made and sealed to an envelope by intervention staff using Statistical Analysis Software. The randomization process was performed by an independent anesthesiologist at the end of the surgical procedure, separated from the ward's nursing staff and outcome assessors to reduce potential bias.

It was impractical to blind the intervention staff in view of the intervening measure, but contrary to the outcome evaluators. To ensure blinding during outcome assessments in the surgical ward, a schedule was established to stagger the intervention and evaluation implementation. Ward staff were informed about the activity monitoring study but were not provided with any specific hypothetical details to prevent potential bias from influencing the study results. Given the intervention characteristics, blinding the patients was also deemed to be unfeasible. To minimize performance bias, patients were informed that we were evaluating 2 mobilization schemes on the understanding that neither scheme was expected to confer any potential advantage.

Data Collection

Preoperative baseline data were collected before an operation as follows: demographic data; preoperative medication and abdominal surgery history, and preoperative radiotherapy and chemotherapy history; vital signs; available preoperative examinations; and preoperative physical activity ability assessed by the International Physical Activity Questionnaire (IPAQ) [25]. Operation-related data collected on the day of surgery included surgical duration, intraoperative bleeding, site, type, and catheter placement status. Postoperative data were collected after each operation as follows: complications according to the Clavien-Dindo classification during hospitalization and 1 week, 2 weeks, and 1 month after the operation; the wearable device ep pod record of the activity time; digital pain Numeric Rating Scale (NRS) scores of rest and exercise; length of hospitalization; time of first exhaust and defecation; and opioids consumption; when discharged from hospital, patients in the intervention group were asked about concerns for feedback on activity time and whether they would supplement activities.

Outcome Measurement

Primary Outcome

The primary outcome was the comprehensive complication index (CCI) [26] within 30 days after surgery. The CCI measured the overall complication calculated for each patient, ranging from 0 (no complication) to 100 (death), including any minor complications instead of focusing on specific ones. It was thus an objective measure with comprehensive and accurate evaluation [27-29]. When calculating CCI, all postoperative complications were evaluated and classified based on the Clavien-Dindo classification, then converted into a weighted continuous scale including quantity and severity. Finally, the CCI value was evaluated using the CCI calculator available [30].

Secondary Outcomes

Daily activity time was set from 12 AM to 11:59 PM based on the data uploaded by ePM/ep pod until the 4th day or until discharge, whichever came first. The patient's compliance was considered to be low [31] on the condition that the patient wore it for less than 8 hours between 8 AM and 8 PM. The degree of pain for rest and exercise in the first 3 days after operation was scored by the NRS [32]. NRS used 0 - 10 points to represent different degrees of pain scored by patients according to their subjective feelings, 0 for no pain and 10 for severe pain. The length of stay was set as the actual length of stay after a patient's operation. The percentage of reaching the scheduled mobilization target was calculated. The patient's activity was considered to be not up to standard if the daily activity time after their operation did not reach the target time. The mortality rate of 30 days after operation was also counted. The first exhaust and defecating times were subject to the patient's own assessment, meaning that an obvious feeling of exhaust or defecation was reported.

Sample Size Calculation

The sample size calculation was based on the primary outcome measured by the CCI 30 days after operation. CCI represents the measurement data that follows a normal distribution, as detailed in the previous trials on postoperative complications [33,34]. The average postoperative CCI of CRC was 10 (SD 5). Assuming that the visual feedback of the movement time could reduce CCI by 25%, the average CCI of the feedback group was set to 7.5 (SD 5) [35]. Based on the above effect size, Power Analysis and Sample Size 2015 Software was used to calculate that a minimum of 172 patients (n=86 per group) would provide 90% power to detect a 25% decrease of CCI and an α of .05 (2-sided test), assuming the feedback group was more favorable compared to the control group [35]. To compensate for a potential 20% dropout rate, 216 patients were required to be included (n=108 per group).

Statistical Analysis

We followed the intention-to-treat principle analyzing all random patients in the baseline analysis. The analysis results of the study were based on the modified intention-to-treat

(MITT) set. Statistical analyses were performed using SPSS version 23.0. The 2-sided significance level was set at a *P* value <.05.

Continuous variables with independent access to mean and SD were initially subjected to the Kolmogorov-Smirnov test to assess their normality. Otherwise, the Shapiro-Wilk test was chosen. The Levene test was used for the analysis of homogeneity of variance. Data with skewed distributions are reported as medians with IQR, while data adhering to a normal distribution are given as means with SD. Categorical variables are expressed as frequencies, percentages, and ratios. For baseline comparisons, normally distributed continuous variables with comparable variance were analyzed using Student *t* test. In contrast, continuous variables with skewed distributions or uneven variance were analyzed with the Mann-Whitney *U* test. Categorical variables were compared using either the chi-square or Fisher exact tests.

The impact of the primary outcome CCI within 30 days after an operation was analyzed using the Student *t* test or the Mann-Whitney *U* test. Except for the MITT analysis, subjects with low device compliance (wearing ePM/ep pod for less than 8 h from 8 AM to 8 PM) were excluded for sensitivity analysis of the primary outcome. In addition, subgroup analyses (preselected age, ASA, preoperative chemoradiotherapy, surgical site, and preoperative mobility) were carried out. Simultaneously, we examined the interaction between the treatment effect and these 5 subgroup variables. In addition, we statistically described the distribution of postoperative complications at all levels in the 2 groups and performed a chi-square test between the groups. As a post hoc analysis, the Spearman/point-biserial correlation coefficient was used to determine the correlation between CCI and the postoperative movement time. In the preliminary analysis, the linear regression model showed a good fitting effect. Based on this result, we conducted a linear regression analysis to assess the relationship between the movement time of POD3 (for patients discharged on POD3, the activity time on POD2 was used to fill in the data) and CCI, adjusted for potential factors such as age, ASA, sex, preoperative chemoradiotherapy, and BMI, all factors which could influence postoperative recovery.

In analyses of secondary outcomes, the Mann-Whitney *U* or Student *t* tests were used to analyze the other measurement data except for repeated measurements. The chi-square test was used to analyze the percentage of patients who achieved the predetermined mobilization goal. Binary logistic regression analysis was used to determine the determinants of patient cohorts to achieve mobilization goals. The regression model considered the following variables: age, sex, ASA grade, surgical site, and the IPAQ [25]. Repeated measurement of the NRS pain score, using a generalized linear mixed model with multinomial distribution, was used for within-group comparisons.

Results

A total of CRC 239 patients were recruited between February 2023 and September 2023. Of these, 216 were randomized, 108 into usual care and 108 with activity feedback. Two patients who met the exclusion criteria were inadvertently enrolled due to unreported medical histories and were subsequently excluded from the MITT analysis (Figure 1). Patients in each randomized group had similar baseline and perioperative characteristics (Table 1).

As shown in Table 2, no significant differences were found in CCI within 30 postoperative days between the control group, with a median CCI of 0 (range, 0 - 20.90), and the activity feedback group, with a median CCI of 0 (range, 0 - 12.20). The estimated mean difference between the 2 groups was −0.59 (95% CI −3.56 to 2.38; *P*=.66). Removing 15 patients with low device compliance as a sensitivity analysis, the estimated mean difference was −0.98 (95% CI −4.07 to 2.10; *P*=.54). The subgroup analysis results revealed no significant interaction between the two groups on CCI within 30 days and all the subgroup variables (Figure 2 and Checklist 1).

Figure 1. Flowchart showing patient inclusion and exclusion.

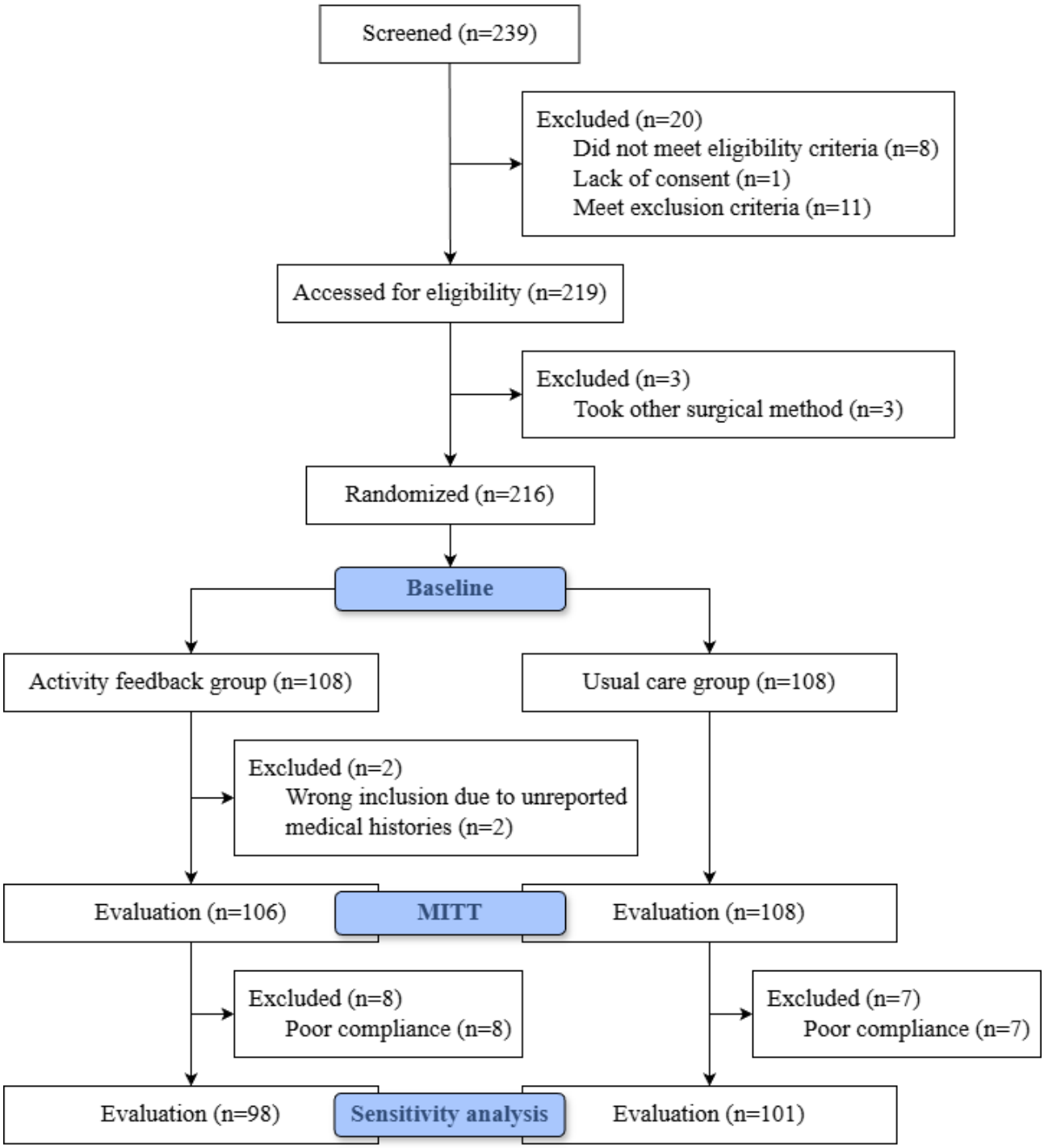


Table . Baseline and perioperative characteristics.

Patient characteristics	Usual care (n=108)	Activity feedback (n=108)	Mann-Whitney *U* test	Chi-square (*df*)	*t* test (*df*)	ρ	Continuity correction factor	*P* value
Age (y)	60.5 (54.25 - 68.75)	63 (55-68)	5998.5	—	—	—	—	.72
Sex, n (%)			—	0.020 (1)	—	—	—	.89
Male	67 (62)	68 (62)						
Female	41 (38)	40 (38)						
BMI[a] (kg/m²), mean (SD)	23.5 (3.3)	23.6 (3.2)	—	—	0.125 (214)	—	—	.90
ASA[b] grade, n (%)			—	—	—	0.016	—	.82
I	1 (0.9)	2 (1.9)						
II	97 (89.8)	94 (87.0)						
III	10 (9.3)	12 (11.1)						
Smoking history, n (%)	22 (20.4)	19 (17.6)	—	0.271 (1)	—	—	—	.60
POC[c], n (%)	46 (42.6)	44 (40.7)	—	0.076 (1)	—	—	—	.78
Hypertension, n (%)	34 (31.5)	33 (30.6)	—	0.022 (1)	—	—	—	.88
Diabetes, n (%)	20 (18.5)	12 (11.1)	—	2.348 (1)	—	—	—	.13
CHD[d], n (%)	4 (3.7)	9 (8.3)	—	2.046 (1)	—	—	—	.15
Preoperative chemoradiotherapy, n (%)	15 (13.9)	12 (11.1)	—	0.381 (1)	—	—	—	.54
IPAQ[e], n (%)			—	—	—	0.053	—	.43
High	23 (21.3)	20 (18.5)						
Medium	69 (63.9)	68 (63.0)						
Low	16 (14.8)	20 (18.5)						
Surgical site, n (%)			—	1.506 (1)	—	—	—	.22
Colon	46 (42.6)	55 (50.9)						
Rectum	62 (57.4)	53 (49.1)						
Laparoscopic, n (%)	103 (95.4)	105 (97.2%)	—	—	—	—	0.130	.72
Converted, n (%)	1 (0.9)	0	—	—	—	—	0.000	>.99
Surgery duration (min), median (IQR)	150 (125 - 183.75)	155 (130-185)	6031.5	—	—	—	—	.66
Bleeding volume (ml), median (IQR)	50 (20-50)	50 (20-50)	5989.5	—	—	—	—	.71
POEM[f] (mg), mean (SD)	56.3 (27.9)	58.1 (31.3)	—	—	0.446 (214)	—	—	.66

[a]BMI: Body Mass Index.

[b]ASA: American Society of Anesthesiologists.

[c]POC: preoperative complications.

[d]CHD: coronary heart disease.

[e]IPAQ: International Physical Activity Questionnaire.

[f]POEM: postoperative equivalent morphine consumption. Sufentanil 0.01 mg equal to morphine 10 mg; Hydromorphone 1.5 mg equal to morphine 10 mg; Nalbuphine 10 mg equal to morphine 10 mg [36].

Table . Between-group difference in CCI[a] 30 days after operation.

	Usual care (n=108)	Activity feedback (n=106)	Mean difference	*P* value
CCI, median (IQR)	0 (0 to 20.90)	0 (0 to 12.20)	−0.59 (−3.56 to 2.38)	.66
CCI[b], median (IQR)	0 (0 to 20.90)	0 (0 to 14.38)	−0.98 (−4.07 to 2.10)	.54
Maximum complication level, n (%)				
0	63 (58.3)	64 (60.4)	—[c]	—
1	14 (13)	17 (16)	—	—
2	26 (24.1)	21 (19.8)	—	—
3[b]	0	1 (0.9)	—	—
3[d]	5 (4.6)	1 (0.9)	—	—
4[b]	0	1 (0.9)	—	—
4[d]	0	1 (0.9)	—	—
5 (death)	0	0	—	—
Severe complications[e], n (%)	5 (4.6)	4 (3.8)	—	—

[a]CCI: comprehensive complications index.

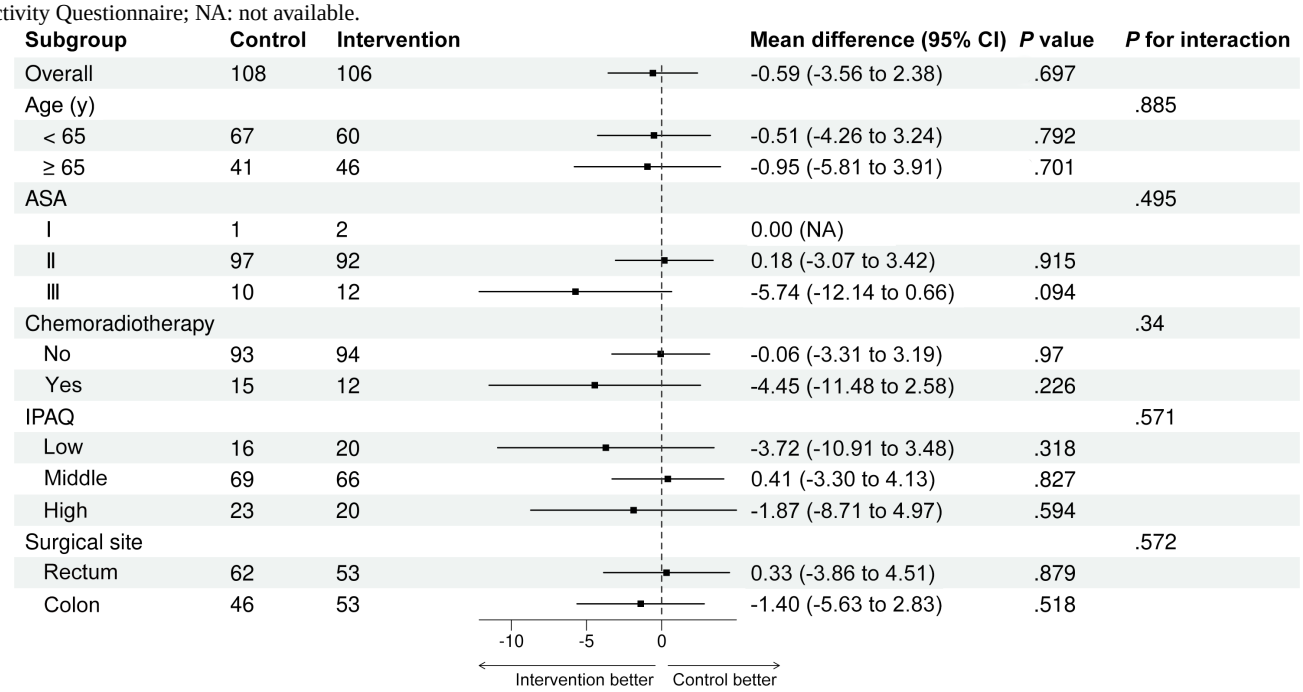
[b]Removing 15 patients with low device compliance.

[c]Not available.

[d]CCI grading was according to the classification of Clavien-Dindo classification.

[e]Severe complication defined as complication ≥3 [35].

Figure 2. Subgroup analysis forest plot for primary outcome indicators. ASA: American Society of Anesthesiologists; IPAQ: International Physical Activity Questionnaire; NA: not available.



The rate of no complications accounted for 58.3% (63/108) of patients in the usual care group and 60.4% (64/106) in the activity feedback group, while severe complications accounted for 4.6% (5/108) and 3.8% (4/106), respectively (Table 2). Both groups had the highest proportion of grade 2 complications. In addition, the incidence of urinary retention was the highest (15/108, 13.9% in the usual care group and 11/106, 10.4% in the activity feedback group) among the operative complications. The incidence rates of complications of infectious, respiratory system, and cardiovascular systems were 6.5% (7/108) vs 9.4% (11/106), 5.5% (6/108) vs 2.8% (3/106), and 1.9% (1/108) vs 1.8% (2/106), respectively, in the usual care group and in the activity feedback group (Table 3). On the analysis of specific complications, by the Fisher exact test, the incidence rate for pneumonia in the routine care group was 4.6% (5/108) and that in the activity feedback group was 1.9% (2/106), *P*=.45; the incidence rate for ileus in the routine care group was 0.9% (1/108) and that in the activity feedback group was 3.8% (4/106), *P*=.21, which indicated that the intervention had no significant effect on the incidence of these specific complications.

Table . Type of complications 30 days after operation.

	Usual care (n=108), n (%)	Activity feedback (n=106), n (%)
Cardiovascular system		
Acute myocardial infarction	0 (0)	1 (0.9)
Arrhythmia	2 (1.9)	1 (0.9)
Respiratory system		
Pneumonia	5 (4.6)	2 (1.9)
Atelectasis	1 (0.9)	0 (0)
Pleural effusion	0	1 (0.9)
Infectious complications		
Wound infection	3 (2.8)	6 (5.7)
Intraperitoneal or retroperitoneal abscess	0	2 (1.9)
Other infections[a]	4 (3.7)	3 (2.8)
Operative complication		
Anastomotic leakage	6 (5.6)	5 (4.7)
Ileus	1 (0.9)	4 (3.8)
Hemorrhage	0 (0)	3 (2.8)
Anemia	8 (7.4)	4 (3.8)
Urinary retention	15 (13.9)	11 (10.4)
Fat liquefaction	3 (2.8)	2 (1.9)
Other gastrointestinal complications[b]	6 (5.6)	7 (6.6)

[a]Other infections: including leukocytosis (white blood cells) >12 and temperature >38°C, shingles.

[b]Other gastrointestinal complications, including diarrhea, constipation, and gastroparesis.

As shown in Table 4, the 2 groups had no significant difference in daily activity time. Similarly, there was no difference in the rate of reaching the target between groups. The usual care and activity feedback groups demonstrated no significant difference in the length of postoperative hospital stay, with a median of 4 days for both groups (*P*=.99), in the first exhaust time (37.13, IQR 22.73-58.77 h vs 38.29, IQR 22.15-57.49 h, respectively, in the usual care group and in the activity feedback group; *P*=.89), and in the first defecating time (44.50, IQR 24.04-68.88 h vs 44.13, IQR 25.81-67.44 h; *P*=.91, respectively, in the usual care group and in the activity feedback group; Table 5). The results of the generalized linear mixed model (Table 6) revealed that the NRS scores of rest and activity showed a declining trend over time, with a significant time effect that was independent of the patient groups (*P*<.001). Moreover, there were no significant interactions between the intervention and the NRS scores at rest (*P*=.54) or during movement (*P*=.89).

In post hoc analysis, there was a significant negative correlation between CCI within the 30th postoperative day and the activity duration on POD2 (r=−0.166; *P*=.02) and POD3 (r=−0.264; *P*=.002). A linear regression analysis showed that the movement time of POD3 was negatively correlated with CCI within the 30th postoperative day, and the standardized β was −.025 (95% CI −.045 to −.006; *P*=.01). Additionally, the longer the activity time of POD3 before 215 minutes, the faster the CCI decreased; however, beyond 215 minutes, the rate of the decline slowed (Figure 3).

Table . Daily activity time and the rate of reaching the target.

	Usual care	Activity feedback	Mean difference/odds ratio (95% CI)	*P* value
Activity time, min				
POD1[a], median (IQR)	105 (74 to 133) (n=107)	114 (65 to 131) (n=105)	−0.7 (−13.4 to 12.0)[b]	.89
POD2[c], median (IQR)	174 (117 to 225) (n=103)	183 (134.5 to 229.5) (n=101)	5.2 (−13.1 to 23.5)[b]	.58
POD3[d], median (IQR)	201 (132 to 248) (n=71)	195 (125.8 to 260.5) (n=70)	−3.8 (−31.8 to 24.1)[b]	.73
Reaching the target				
POD1, n/N (%)	40/107 (37.4)	46/105 (43.8)	0.8 (0.4 to 1.4)[e]	.34
POD2, n/N (%)	20/103 (19.4)	21/101 (20.8)	1.1 (0.5 to 2.2)[e]	.81
POD3, n/N (%)	4/71 (5.6)	2/70 (2.9)	0.5 (0.1 to 3.0)[e]	.69

[a]POD1: the first day after operation.
[b]Mean difference (95% CI).
[c]POD2: the second day after operation.
[d]POD3: the third day after operation.
[e]Odds ratio (95% CI).

Table . Postoperative recovery index.

	Usual care (n=108), median (IQR)	Activity feedback (n=106), median (IQR)	Mean difference (95% CI)	*P* value
LOS[a] (d)	4 (3 to 5)	4 (4 to 5)	−0.4 (−1.49 to 0.69)	.47
First exhaust time (h)	37.13 (22.73 to 58.77)	38.29 (22.15 to 57.49)	0.13 (−8.07 to 8.34)	.98
First defecating time (h)	44.50 (24.04 to 68.88)	45.13 (25.81 to 67.44)	−2.99 (−14.00 to 8.03)	.60

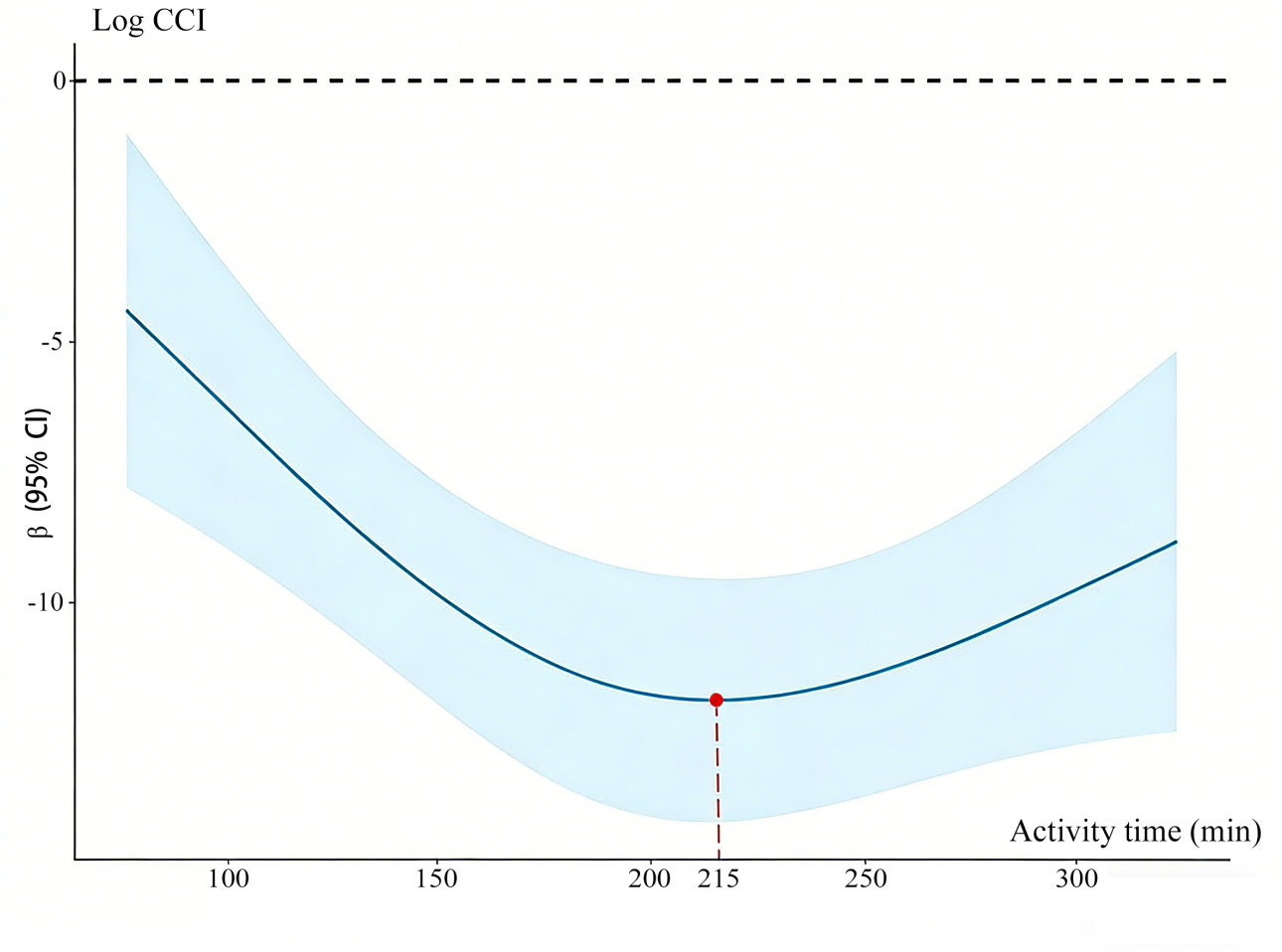
[a]LOS: length of postoperative hospital stay.

Table . Pain scores after surgery at rest or movement.

Pain scores	Usual care (n=108), median (IQR)	Activity feedback (n=106), median (IQR)	Fixed effect *P* value		
			Group	Time	Group*time
NRS[a] at rest			.95	.007	.54
POD1[b]	1 (0 - 2)	1 (0 - 2)			
POD2[c]	0 (0 - 2)	0 (0 - 2)			
POD3[d]	1 (0 - 1.75)	1 (0 - 1)			
NRS at movement			.32	.001	.89
POD1	2 (1 - 4)	2 (2 - 4)			
POD2	2 (1 - 3.75)	2 (1 - 4)			
POD3	2.5 (2 - 3)	2.5 (1.75 - 3)			

[a]NRS: Numeric Rating Scale.
[b]POD1: the first day after operation.
[c]POD2: the second day after operation.
[d]POD3: the third day after operation.

Figure 3. Multivariate restricted cubic spline analysis of activity time and comprehensive complications index (CCI).



Discussion

Principal Findings

In this study, the activity incentive feedback based on a wearable device (ePM/ep pod) did not reduce the CCI of patients with CRC within postoperative 30 days. The differences between groups in terms of daily activity time, length of postoperative hospital stay, intestinal function recovery, and NRS pain scores were also not significant.

There may be the following reasons for the lack of difference in CCI and daily activity durations between the 2 groups. First, this study was conducted in a surgical ward with relatively deepening ERAS guidelines, where preoperative education and routine activity supervision may have facilitated effective postoperative activities even in the absence of the feedback from wearable devices. Second, all the subjects in the present study received the same activity encouragement by medical staff according to the routine care and were set the same daily activity target; therefore, patients in the control group with high compliance might obtain activity feedback through other means. Third, laparoscopy accounted for more than 95% of CRC surgery in the present study, which was similar to the type of surgery used in another study that yielded negative results regarding the hypothesis that early postoperative activity could reduce complications [15]. Future studies should investigate whether such incentive feedback would be effective for patients experiencing more intense surgical stimulation or possessing less ERAS education. Previous studies using step count feedback for mobility-enhancing showed conflicting

The assessment of early postoperative activity achievement in patients with CRC directly reflects the practical outcomes of the intervention strategies employed in this study. The activity compliance rate of patients with CRC on POD1 was 37.4% (in the usual care group) and 43.8% (in the activity monitoring feedback group), which was much lower than expected, indicating that even with the feedback mechanism, there was only a marginal improvement in patients' adherence to the activity targets. Based on the current 6 hours recommended by ERAS guidelines, this study set a less active target of 2 hours for POD 0. In addition, the maximum activity time on POD1 was 310 minutes, which was much lower than the 6 hours recommended by ERAS guidelines. The relatively low compliance rates suggest that factors beyond the provision of activity feedback may be at play in influencing patients' postoperative behavior, such as the intensity and appropriateness of the activity targets set. The insights gained from analyzing the achievement of early postoperative activity not only provide a deeper understanding of the limitations of the feedback intervention but also highlight the need for more tailored and patient-centered approaches to enhance postoperative mobilization. The postoperative activity time was measured by an electronic parameter objectively recorded by a wearable device, and the measurement method may be different from the

considerations of previous guidelines. This may be the reason why the postoperative activity time was inconsistent with most previous studies.

Previous studies [37,38] have found patients undergoing major visceral (abdominal) surgery may not be able to achieve the mobilization goals in the early postoperative period. An observational pilot study [39] pointed out that only 50% of patients undergoing major visceral surgery could walk until POD2. The cumulative duration of postoperative activity per day ranged from 15 to 155 minutes, and only 40% of patients were satisfied with activity or physical exercise. There is a discrepancy between ERAS goals and clinical practice. Another study [40] reported that only a few patients (23.5%) achieved the mobilization goal of 6 hours (ERAS guidelines) of activity on the first day after CRC surgery. Ramírez et al [41] set the early activity goal for patients with CRC as sitting for at least 6 hours on the first postoperative day. However, the achievement rate was only 44.6%. In general, the mobilization goals described in most studies were either unclearly defined, imprecisely measured, or inaccurately monitored.

In addition, the early activity goals in ERAS did not differentiate between disease types, and radical abdominal tumor surgery is far more invasive than other simple abdominal organ partial removal. Patients with tumors were often elderly and had comorbidities, which can limit physical fitness and motivation. A study by Reed et al [42] on activity tracking after bariatric surgery found a significant negative correlation between age and the steps on POD1. Our study presented similar findings in that age was an independent factor affecting the achievement of activity on POD1. The older the patients were, the lower the possibility of reaching the activity goal. Furthermore, age negatively affected activity duration on POD2 and POD3. All patients with CRC in the present study were 61 years on average, and 42% had preoperative comorbidities, which may also be the reason for the low level of postoperative activity. Also, regression analysis showed that the activity duration of patients with colon cancer on POD2 was significantly longer than that of patients with rectal cancer, being 26.42 minutes longer. Patients with colon cancer were 2.28 times more likely than those with rectal cancer to meet the activity standard on POD2. It can be seen that the mobility of patients with colon cancer on POD2 was higher than that of patients with rectal cancer. We speculate that the reason for this finding may be that colon cancer basically does not receive radiotherapy, the proportion of preoperative chemotherapy may be lower than for rectal cancer, and the proportion of temporary ileostomy of colon cancer being much lower than that for rectal cancer. Our finding also suggests the target of early activity recommended by current ERAS guidelines (6 h in POD1) may be too high for these patients. A further increase in activity did not increase the benefit when activity levels reached a certain level. Considering our result of the regression analysis, 215 minutes might be considered as the target for activity time on POD3 to decrease CCI in patients with CRC.

In addition, several studies [43-45] have reported that early postoperative mobilization (within 24 h) was associated with fewer postoperative complications, fewer major complications, faster intestinal function recovery, and shorter hospital stays. The regression analysis in this study found that patients with high levels of physical activity before surgery were about 27 minutes more active the next day compared to patients with moderate levels of physical activity before surgery, and about 37 minutes more active than patients with low levels of physical activity before surgery. Therefore, perhaps by exploring the extent of postoperative activity decline caused by different surgical trauma, the approximate postoperative activity amount of the patient can be inferred and then combined with the preoperative activity level of the patient, a personalized postoperative activity plan can be formulated.

Limitations

This study had some limitations. First, the 216 patients were all admitted to one hospital; therefore, the conclusion still needs to be verified in more hospitals where ERAS guidelines are less practiced. The low proportion of ASA 3 patients in this study may limit our comprehensive assessment of the benefits of patients with high comorbidities from feedback from postoperative activities. Future research may consider including more ASA3 patients to better understand the response of this special population to postoperative activity feedback intervention. Second, since it was not possible to blind the patients in this study, the 2 groups of patients may have influenced each other. Due to practical difficulties such as high patient mobility and limited follow-up time, this study failed to collect information about the views of patients and clinicians with the wearable device and how motivated they felt by wearing this. In the future, the study can be expanded to multiple wards to achieve cluster randomization, which may reduce bias. Future research can also be combined with qualitative items to provide a more comprehensive understanding of patient and clinician experiences and perceptions of wearable devices. Third, as 96% of patients in this study underwent laparoscopic surgery, the results may not be generalized to open surgery or other types of surgery. Although our inclusion criteria were broad, we excluded patients who had difficulty with preoperative activity or were in a poor physical condition. These patients may benefit from the incentive of activity feedback. A final limitation was there may be a difference between self-reported preoperative mobility assessed by IPAQ, which was not designed for hospital patients, but for assessing physical activity in daily life, and actual quantitative mobility.

Conclusions

In the context of ERAS, this randomized controlled trial did not find evidence that activity stimulation based on feedback from ePM/ep pod wearable devices could reduce 30-day postoperative CCI in patients who underwent CRC surgery. However, ePM/ep pod could accurately record daily activity duration, which may be negatively correlated with CCI on POD3.

Acknowledgments

The authors wish to thank Chong Lei and Ziyu Zheng for participating in protocol discussion.

Funding

This research is supported by the University-funded clinical research project (2022L2213) and Hospital-funded military medical research project (JSYXZ12) to HN.

Data Availability

The datasets generated or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

Authors' Contributions

HN and HD designed the study. YM, YW, and HW conducted the study. NY and ZZ performed the data collection and data analysis. FF wrote the original draft of the paper. YM and HN contributed to critical revision. FF and HN reviewed and edited the manuscript and have approved its final version. HN is responsible for the overall content as guarantor.
HD is the co-corresponding author (email: hldong6@hotmail.com; phone: 8613669226699).

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH (V 1.6.1) checklist [46].
[PDF File, 11188 KB - mhealth_v14i1e70534_app1.pdf]

References

1. Zygulska AL, Pierzchalski P. Novel diagnostic biomarkers in colorectal cancer. Int J Mol Sci 2022 Jan 13;23(2):852. [doi: 10.3390/ijms23020852] [Medline: 35055034]
2. Baidoun F, Elshiwy K, Elkeraie Y, et al. Colorectal cancer epidemiology: recent trends and impact on outcomes. Curr Drug Targets 2021;22(9):998-1009. [doi: 10.2174/1389450121999201117115717] [Medline: 33208072]
3. van Rooijen S, Carli F, Dalton S, et al. Multimodal prehabilitation in colorectal cancer patients to improve functional capacity and reduce postoperative complications: the first international randomized controlled trial for multimodal prehabilitation. BMC Cancer 2019 Jan 22;19(1):98. [doi: 10.1186/s12885-018-5232-6] [Medline: 30670009]
4. Namba Y, Hirata Y, Mukai S, et al. Clinical indicators for the incidence of postoperative ileus after elective surgery for colorectal cancer. BMC Surg 2021 Feb 11;21(1):80. [doi: 10.1186/s12893-021-01093-7] [Medline: 33573636]
5. Khuri SF, Henderson WG, DePalma RG, et al. Determinants of long-term survival after major surgery and the adverse effect of postoperative complications. Ann Surg 2005 Sep;242(3):326-341. [doi: 10.1097/01.sla.0000179621.33268.83] [Medline: 16135919]
6. Aoyama T, Oba K, Honda M, et al. Impact of postoperative complications on the colorectal cancer survival and recurrence: analyses of pooled individual patients' data from three large phase III randomized trials. Cancer Med 2017 Jul;6(7):1573-1580. [doi: 10.1002/cam4.1126] [Medline: 28639738]
7. Arnarson Ö, Butt-Tuna S, Syk I. Postoperative complications following colonic resection for cancer are associated with impaired long-term survival. Colorectal Dis 2019 Jul;21(7):805-815. [doi: 10.1111/codi.14613] [Medline: 30884061]
8. Schwenk W. Optimized perioperative management (fast-track, ERAS) to enhance postoperative recovery in elective colorectal surgery. GMS Hyg Infect Control 2022;17:Doc10. [doi: 10.3205/dgkh000413] [Medline: 35909653]
9. Portinari M, Ascanelli S, Targa S, et al. Impact of a colorectal enhanced recovery program implementation on clinical outcomes and institutional costs: a prospective cohort study with retrospective control. Int J Surg 2018 May;53:206-213. [doi: 10.1016/j.ijsu.2018.03.005] [Medline: 29548700]
10. Chinese Society of Surgery, Chinese Society of Anesthesiology. Consensus on ERAS and guidelines for pathway management in China (2018). Chin J Pract Surg 2018;38(1):1-20. [doi: 10.19538/j.cjps.issn1005-2208.2018.01.01]
11. Gustafsson UO, Scott MJ, Hubner M, et al. Guidelines for perioperative care in elective colorectal surgery: enhanced recovery after surgery (ERAS®) society recommendations: 2018. World J Surg 2019 Mar;43(3):659-695. [doi: 10.1007/s00268-018-4844-y] [Medline: 30426190]
12. Xiang B, Jiao S, Si Y, Yao Y, Yuan F, Chen R. Risk factors for postoperative pneumonia: a case-control study. Front Public Health 2022;10:913897. [doi: 10.3389/fpubh.2022.913897] [Medline: 35875004]
13. Li Q, Du L, Lu L, et al. Clinical application of enhanced recovery after surgery in perioperative period of laparoscopic colorectal cancer surgery. J Laparoendosc Adv Surg Tech A 2019 Feb;29(2):178-183. [doi: 10.1089/lap.2018.0708] [Medline: 30614769]

14. Gustafsson UO, Oppelstrup H, Thorell A, Nygren J, Ljungqvist O. Adherence to the ERAS protocol is associated with 5-year survival after colorectal cancer surgery: a retrospective cohort study. World J Surg 2016 Jul;40(7):1741-1747. [doi: 10.1007/s00268-016-3460-y] [Medline: 26913728]
15. Fiore JF Jr, Castelino T, Pecorelli N, et al. Ensuring early mobilization within an enhanced recovery program for colorectal surgery: a randomized controlled trial. Ann Surg 2017 Aug;266(2):223-231. [doi: 10.1097/SLA.0000000000002114] [Medline: 27997472]
16. Grass F, Pache B, Martin D, et al. Feasibility of early postoperative mobilisation after colorectal surgery: a retrospective cohort study. Int J Surg 2018 Aug;56:161-166. [doi: 10.1016/j.ijsu.2018.06.024] [Medline: 29935366]
17. Tazreean R, Nelson G, Twomey R. Early mobilization in enhanced recovery after surgery pathways: current evidence and recent advancements. J Comp Eff Res 2022 Feb;11(2):121-129. [doi: 10.2217/cer-2021-0258] [Medline: 35045757]
18. Twomey R, Culos-Reed SN, Daun JT, Ferber R, Dort JC. Wearable activity trackers and mobilization after major head and neck cancer surgery: you can't improve what you don't measure. Int J Surg 2020 Dec;84:120-124. [doi: 10.1016/j.ijsu.2020.10.032] [Medline: 33157275]
19. Wiklund M, Sundqvist E, Fagevik Olsén M. Physical activity in the immediate postoperative phase in patients undergoing Roux-en-Y gastric bypass-a randomized controlled trial. Obes Surg 2015 Dec;25(12):2245-2250. [doi: 10.1007/s11695-015-1690-y] [Medline: 25910983]
20. No JH, Kim K, Kim YB, et al. Effects of an activity tracker with feedback on physical activity in women after midline laparotomy: a randomized controlled trial. J Obstet Gynaecol Res 2021 Jul;47(7):2544-2550. [doi: 10.1111/jog.14807] [Medline: 33899302]
21. Keats MR, Yu X, Sweeney Magee M, et al. Use of wearable activity-monitoring technologies to promote physical activity in cancer survivors: challenges and opportunities for improved cancer care. Int J Environ Res Public Health 2023 Mar 8;20(6):4784. [doi: 10.3390/ijerph20064784] [Medline: 36981693]
22. Singh B, Zopf EM, Howden EJ. Effect and feasibility of wearable physical activity trackers and pedometers for increasing physical activity and improving health outcomes in cancer survivors: a systematic review and meta-analysis. J Sport Health Sci 2022 Mar;11(2):184-193. [doi: 10.1016/j.jshs.2021.07.008] [Medline: 34314878]
23. Ma LN, Nie H, Ji G, et al. Accuracy of wearable device in monitoring postoperative motion and sleep of patients. Chin J Anesthesiol 2020;40(10):1232-1236. [doi: 10.3760/cma.j.cn131073.20200820.01020]
24. The Ethics Committee of the First Affiliated Hospital of the Air Force Military Medical University. Medical ethics committee charter. URL: https://www.fmmu.edu.cn/info/1404/4453.htm [accessed 2026-01-22]
25. Roberts-Lewis SF, White CM, Ashworth M, Rose MR. The validity of the international physical activity questionnaire (IPAQ) for adults with progressive muscle diseases. Disabil Rehabil 2022 Nov;44(23):7312-7320. [doi: 10.1080/09638288.2021.1983042] [Medline: 34606392]
26. Slankamenac K, Graf R, Barkun J, Puhan MA, Clavien PA. The comprehensive complication index: a novel continuous scale to measure surgical morbidity. Ann Surg 2013 Jul;258(1):1-7. [doi: 10.1097/SLA.0b013e318296c732] [Medline: 23728278]
27. Kudo T, Oshikiri T, Goto H, et al. Comprehensive complication index as a prognostic factor in minimally invasive esophagectomy for esophageal squamous cell carcinoma. Esophagus 2022 Jul;19(3):410-416. [doi: 10.1007/s10388-022-00911-y] [Medline: 35220510]
28. Artiles-Armas M, Roque-Castellano C, Conde-Martel A, Marchena-Gómez J. The comprehensive complication index is related to frailty in elderly surgical patients. J Surg Res 2019 Dec;244:218-224. [doi: 10.1016/j.jss.2019.06.011] [Medline: 31301477]
29. Kim SH, Hwang HK, Lee WJ, Kang CM. Comprehensive complication index or Clavien-Dindo classification: which is better for evaluating the severity of postoperative complications following pancreatectomy? World J Surg 2021 Mar;45(3):849-856. [doi: 10.1007/s00268-020-05859-7] [Medline: 33191470]
30. Home of AssesSurgery. AssesSurgery GmbH. URL: https://www.assessurgery.com/ [accessed 2026-01-21]
31. Finet M, Bellicha A, Sage E, et al. Comprehensive assessment of postoperative mobility during the first days after mini-invasive lung surgery: a prospective observational study. J Clin Anesth 2023 Jun;86:111048. [doi: 10.1016/j.jclinane.2022.111048] [Medline: 36716650]
32. Gagliese L, Weizblit N, Ellis W, Chan VWS. The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. Pain 2005 Oct;117(3):412-420. [doi: 10.1016/j.pain.2005.07.004] [Medline: 16153776]
33. Spanjersberg WR, Reurings J, Keus F, van Laarhoven CJ. Fast track surgery versus conventional recovery strategies for colorectal surgery. Cochrane Database Syst Rev 2011 Feb 16(2):CD007635. [doi: 10.1002/14651858.CD007635.pub2] [Medline: 21328298]
34. Lv L, Shao YF, Zhou YB. The enhanced recovery after surgery (ERAS) pathway for patients undergoing colorectal surgery: an update of meta-analysis of randomized controlled trials. Int J Colorectal Dis 2012 Dec;27(12):1549-1554. [doi: 10.1007/s00384-012-1577-5] [Medline: 23001161]
35. Gloor S, Misirlic M, Frei-Lanter C, et al. Prehabilitation in patients undergoing colorectal surgery fails to confer reduction in overall morbidity: results of a single-center, blinded, randomized controlled trial. Langenbecks Arch Surg 2022 May;407(3):897-907. [doi: 10.1007/s00423-022-02449-0] [Medline: 35084526]

36. Chinese Society of Anesthesiology. Expert consensus on postoperative pain management in adults (2017 Edition). J Clin Anesthesiol 2017;33(9):911-917.
37. van Zelm R, Coeckelberghs E, Sermeus W, et al. Variation in care for surgical patients with colorectal cancer: protocol adherence in 12 European hospitals. Int J Colorectal Dis 2017 Oct;32(10):1471-1478. [doi: 10.1007/s00384-017-2863-z] [Medline: 28717841]
38. He Y, Yang J, Jiang L. Analysis of associated factors of early mobilization in colorectal cancer patients who underwent enhanced recovery after surgery pathway. Chin J Bases Clin Gen Surg 2021;28(12):1599-1603. [doi: 10.7507/1007-9424.202106124]
39. Wolk S, Distler M, Müssle B, Söthje S, Weitz J, Welsch T. Adherence to ERAS elements in major visceral surgery-an observational pilot study. Langenbecks Arch Surg 2016 May;401(3):349-356. [doi: 10.1007/s00423-016-1407-2] [Medline: 27013325]
40. Gustafsson UO, Hausel J, Thorell A, et al. Adherence to the enhanced recovery after surgery protocol and outcomes after colorectal cancer surgery. Arch Surg 2011 May;146(5):571-577. [doi: 10.1001/archsurg.2010.309] [Medline: 21242424]
41. Ramírez JM, Blasco JA, Roig JV, et al. Enhanced recovery in colorectal surgery: a multicentre study. BMC Surg 2011 Apr 14;11:9. [doi: 10.1186/1471-2482-11-9] [Medline: 21489315]
42. Reed B, Tabone LE, Tabone JK, Szoka N, Abunnaja S, Bailey K. The use of an activity tracker to objectively measure inpatient activity after bariatric surgery. Surg Obes Relat Dis 2021 Jan;17(1):90-95. [doi: 10.1016/j.soard.2020.08.033] [Medline: 33032917]
43. Twomey R, Matthews TW, Nakoneshny S, et al. Impact of early mobilization on recovery after major head and neck surgery with free flap reconstruction. Cancers (Basel) 2021 Jun 8;13(12):2852. [doi: 10.3390/cancers13122852] [Medline: 34201003]
44. Ding X, Zhang H, Liu H. Early ambulation and postoperative recovery of patients with lung cancer under thoracoscopic surgery-an observational study. J Cardiothorac Surg 2023 Apr 11;18(1):136. [doi: 10.1186/s13019-023-02263-9] [Medline: 37041603]
45. Cook DJ, Thompson JE, Prinsen SK, Dearani JA, Deschamps C. Functional recovery in the elderly after major surgery: assessment of mobility recovery using wireless technology. Ann Thorac Surg 2013 Sep;96(3):1057-1061. [doi: 10.1016/j.athoracsur.2013.05.092] [Medline: 23992697]
46. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]

Abbreviations

ASA: American Society of Anesthesiologists
CCI: comprehensive complications index
CRC: colorectal cancer
ep pod: electrophysiology pod
ePM: enhanced patient monitoring
ERAS: enhanced recovery after surgery
IPAQ: International Physical Activity Questionnaire
MITT: modified intention-to-treat
NRS: Numeric Rating Scale
POD1: the first day after operation
POD2: the second day after operation
POD3: the third day after operation

Edited by L Buis; submitted 25.Dec.2024; peer-reviewed by A Dosis, G Pisaneschi; revised version received 20.Dec.2025; accepted 28.Dec.2025; published 30.Jan.2026.

Please cite as:
Meng Y, Fan F, Ma Y, Yan N, Wang H, Zhang Z, Wang Y, Dong H, Nie H
Impact of Mobilization Facilitated by Wearable Device Enhanced Patient Monitoring/Electrophysiology Pod–Based Feedback on Postoperative Complications Following Colorectal Cancer Surgery: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e70534
URL: https://mhealth.jmir.org/2026/1/e70534
doi:10.2196/70534

Original Paper

Digital Cognitive Behavioral Therapy for Chronic Insomnia in South Korea: Cost-Effectiveness Analysis Using Decision Tree and Markov Modeling Based on a Secondary Analysis of a Randomized Clinical Trial

Jung Hyun Kim[1*], PhD; Minjeong Sohn[2*], MSc; Jimin Woo[3], MA; Jaeyong Shin[4], MD, PhD; Euna Han[2], PhD

[1]Division of Tourism and Wellness, Hankuk University of Foreign Studies, Yongin, Republic of Korea

[2]College of Pharmacy, Yonsei Institute of Pharmaceutical Sciences, Yonsei Institute for Digital Health, Yonsei University, Incheon, Republic of Korea

[3]AIMNEXT Inc, Seoul, Republic of Korea

[4]College of Medicine, Yonsei University, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Euna Han, PhD
College of Pharmacy
Yonsei Institute of Pharmaceutical Sciences, Yonsei Institute for Digital Health
Yonsei University
Yonsei International Campus, Veritas Hall D, Rm 212
Incheon, 03722
Republic of Korea
Phone: 82 327494511
Email: eunahan@yonsei.ac.kr

Abstract

Background: Insomnia is a prevalent sleep disorder characterized by difficulty initiating or maintaining sleep and is associated with substantial health and economic burdens. Although cognitive behavioral therapy (CBT) is recommended as the first-line treatment, pharmacotherapy remains widely used despite adverse effects and significant indirect costs related to impaired productivity and workplace safety. Digital therapeutics delivering CBT through mobile platforms have emerged as scalable alternatives to improve access and outcomes. Somzz is a commercially available, domestically developed digital therapeutic that delivers CBT-based interventions for insomnia via a mobile app.

Objective: This study evaluated the cost-effectiveness of Somzz compared with conventional insomnia treatment combining CBT and pharmacotherapy from both health care system and societal perspectives in South Korea.

Methods: A decision-analytic model integrating a short-term decision tree with a Markov model was developed to compare costs and outcomes of digital CBT via Somzz versus conventional care in 2023. The model simulated a 27-week time horizon (three treatment cycles) and applied an annual discount rate of 4.5%. Clinical inputs, including remission probabilities and health utility values, were derived from a published randomized clinical trial comparing digital CBT delivered via Somzz with sleep hygiene education. Additional inputs, including health care resource use and unit costs, were obtained from published literature and national sources. Health outcomes were measured in quality-adjusted life years (QALYs). The cost analysis included direct medical costs and indirect costs related to absenteeism, productivity loss, and workplace accidents attributable to insomnia. Incremental cost-effectiveness ratios (ICERs) were estimated in 2023 South Korean Won (KRW). Deterministic one-way and probabilistic sensitivity analyses were conducted to assess uncertainty.

Results: From a health care system perspective, digital CBT via Somzz resulted in modestly higher costs and improved health outcomes compared with standard care. Over approximately 6.5 months, Somzz generated an additional 0.0092 QALYs per patient at an incremental cost of KRW 79,691 (US $61.87), yielding an ICER of KRW 8,719,727 (US $990,883) per QALY gained. This estimate was well below the Korean willingness-to-pay threshold of KRW 30,000,000 (US $23,192.91) per QALY. From a societal perspective, digital CBT was cost-saving, producing a negative ICER due to reductions in health care utilization, workplace accidents, and productivity losses associated with higher remission rates. Sensitivity analyses identified intervention

costs and remission probabilities as key drivers; however, digital CBT remained cost-effective across all scenarios under willingness-to-pay thresholds of KRW 30,000,000 and KRW 15,000,000 per QALY.

Conclusions: Digital CBT for insomnia offers favorable clinical and economic value in South Korea. Using Korean clinical trial data and locally relevant societal cost inputs, this study provides policy-relevant evidence supporting early integration of digital CBT into routine insomnia care, employer health strategies, and national digital health policy.

Trial Registration: World Health Organization International Clinical Trials Registry Platform KCT0007292; https://trialsearch.who.int/Trial2.aspx?TrialID=KCT0007292

(JMIR Mhealth Uhealth 2026;14:e71750) doi:10.2196/71750

KEYWORDS

insomnia; digital therapeutics; cognitive behavioral therapy; cost-effectiveness analysis; QALYs; quality-adjusted life years

Introduction

Problem

Insomnia is a prevalent sleep disorder among adults, characterized by difficulty initiating or maintaining sleep or early morning awakening despite adequate opportunity for sleep, resulting in poor sleep quality and daytime dysfunction [1-3]. Evidence suggests that the global prevalence of insomnia has increased, particularly during the COVID-19 pandemic, with marked rises among health care workers and individuals infected with COVID-19 [4,5]. Insomnia is associated with numerous adverse health outcomes, including a higher risk of psychiatric disorders and diminished health-related quality of life [6-8]. Furthermore, insomnia may lead to decreased work productivity and a higher risk of motor vehicle accidents [9,10]. Consequently, insomnia imposes substantial health and economic burdens, including direct costs related to health care use and medication use, as well as indirect costs from absenteeism and reduced workplace productivity [6,11].

Several treatment guidelines recommend cognitive behavioral therapy (CBT) as a first-line treatment for chronic insomnia [2,12-14]. Pharmacotherapy-centered management, despite its widespread use, is associated with adverse events, limited long-term effects, and continued health care use, while often failing to restore daily functioning and work performance [15,16]. Nevertheless, pharmacotherapy remains the most commonly used treatment in the United States and other countries, including Korea [17-19]. A survey of Korean physicians reported that pharmacotherapy was chosen by 57.14% of patients, followed by sleep hygiene education (37.11%), whereas only 1.18% received CBT [17]. This discrepancy persists despite strong evidence that CBT provides durable benefits that can be maintained for more than 1 year after treatment completion, in contrast to the short-term symptom control typically observed with pharmacotherapy [20]. As a result, indirect societal costs, such as productivity loss and accident-related costs, frequently exceed direct medical expenditures in insomnia [6,21,22].

Review of Relevant Scholarship

Digital cognitive behavioral therapy for insomnia (dCBT-I) has emerged as a scalable solution that can overcome access barriers through the widespread availability of smartphones, tablets, and computers. Recent randomized controlled trials have demonstrated that dCBT-I not only reduces insomnia symptoms but also improves functional health, psychological well-being, and sleep-related quality of life over time [23]. Digital platforms can allow CBT to reach a wide population [24-26]. Meta-analytic evidence robustly supports the efficacy and broader benefits of digital CBT for insomnia. A 2022 meta-analysis demonstrated large short-term improvements in insomnia severity and moderate reductions in depressive symptoms that persisted long-term [27], while another meta-analysis focusing on occupational outcomes found significant decreases in lost productivity (presenteeism) and work-related rumination among working adults [28].

Building on this clinical evidence, recent economic evaluations have also matured: a 2023 systematic review and meta-analysis reported that dCBT-I consistently achieves favorable cost-utility and cost-effectiveness, particularly when societal costs such as productivity losses are considered [29]. Complementary findings from randomized controlled trials show that guided internet-delivered CBT-I can yield measurable quality-adjusted life years (QALY) gains and, from a societal perspective, has a high probability of dominating control conditions [30]. In addition, simulation modeling using a Markov framework in the United States found that fully automated digital CBT programs (eg, Sleepio) may not only be cost-effective but potentially cost-beneficial compared with no treatment, clinician-delivered CBT, and pharmacotherapy [31]. However, to date, no cost-effectiveness analysis of dCBT-I has been conducted using Korean clinical trial data and Korean-specific cost inputs. Given the strong influence of country-specific health care system structures and current treatment patterns on economic outcomes in insomnia, reliance on evidence generated in Western settings with limited consideration of societal costs may substantially misrepresent the value of digital CBT-I in the Korean context [29-32]. This study addresses an important evidence gap by incorporating Korean randomized clinical trial data and locally relevant cost inputs to evaluate the cost-effectiveness of a domestically developed digital CBT-I from both health care system and societal perspectives.

Hypothesis, Aims, and Objectives

Somzz (AIMNEXT Inc) is the first domestically developed digital therapeutic in Korea designed to deliver CBT for insomnia (CBT-I) [33]. The fully automated mobile app incorporates core CBT-I components, including stimulus control, sleep restriction, sleep hygiene education, relaxation training, cognitive therapy, and relapse prevention, while providing

personalized, real-time feedback based on users' sleep diary data and behavioral patterns. By targeting the psychological, behavioral, and cognitive factors that maintain chronic insomnia, Somzz offers a scalable alternative to clinician-delivered CBT-I.

Based on its demonstrated clinical efficacy and the substantial societal burden of untreated insomnia, we hypothesized that digital CBT-I delivered through Somzz would represent a cost-effective strategy compared with conventional care involving CBT and pharmacotherapy. In particular, we anticipated that Somzz would improve health outcomes while reducing overall costs when broader societal impacts, such as productivity losses and accident-related costs, are taken into account.

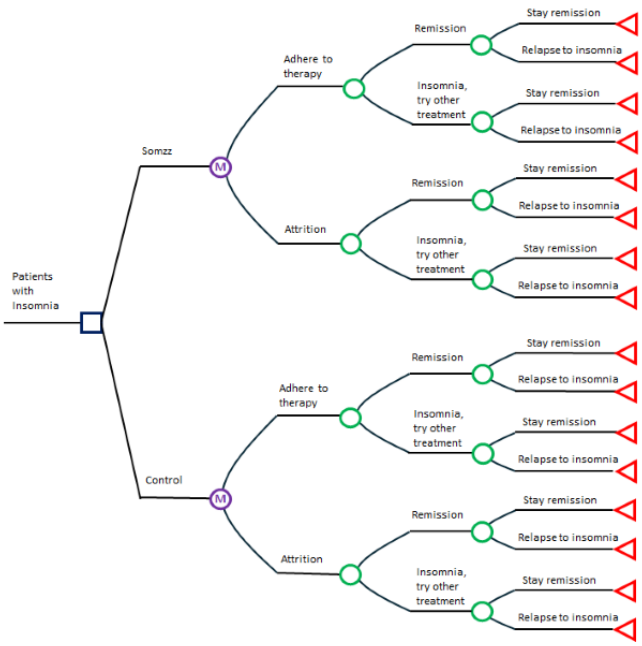
The analysis aimed to evaluate the cost-effectiveness of Somzz compared to standard care from both health care system and societal perspectives by using clinical outcomes from a Korean randomized controlled trial (RCT) and integrating real-world Korean wage, service use, and epidemiological data. We also aimed to examine the robustness of these findings through deterministic and probabilistic sensitivity analyses.

Methods

Conditions and Design

This study is a secondary analysis of a parent single-blind RCT (trial registration: KCT0007292; registered retrospectively on 2022-05-17) to assess the safety and efficacy of a digital therapeutic, Somzz [33]. This study is to compare the costs and outcomes of Somzz versus conventional care in 2023, and thus, we constructed a static decision tree model combined with a Markov model with a 27-week (3 treatment cycles over 9 weeks), which aligns with previous insomnia cost-effectiveness studies [31]. A Markov model represents disease progression by simulating transitions between mutually exclusive health states over discrete time cycles, assuming that future transitions depend only on the current state and not the path taken to reach it [34]. In our model, patients transitioned through health states of adherence, attrition, remission, relapse, and persistence of insomnia (Figure 1).

Figure 1. A basic description of the structure of the decision analytic Markov model.



We adopted the health care system and societal perspectives according to the economic evaluation guidelines in Korea [35]. This dual approach allowed us to assess the economic implications of a health care intervention more thoroughly by capturing the direct financial impact on the health care system and broader economic and societal effects. A discount of 4.5% was applied as per Korean guidelines [35]. We referred to the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) guidelines [36] to conduct the overall cost-effectiveness analysis (Supplementary Table S1 in Multimedia Appendix 1).

Inclusion and Exclusion Criteria

In the aforementioned parent trial, participants were eligible if they were (1) ≥19 years old, (2) had a diagnosis of chronic insomnia, and (3) were able to participate in scheduled visits and assessments. Exclusion criteria in the original trial included: (1) severe psychiatric or neurological conditions, (2) participation in other sleep-related interventions, and (3) medical contraindications to CBT-I or study procedures [33]. This study included only participants from the original trial dataset; no additional inclusion or exclusion criteria were applied.

Participant Characteristics

The parent RCT enrolled 98 adults (Somzz: 49; control: 49). Key characteristics were (1) age: adults 19 years and older (details reported in original trial [33]), (2) employment: 51% engaged in paid work (Somzz: 22/49; control: 28/49), and (3) baseline insomnia severity: clinically significant chronic insomnia. Participants were similar between groups due to randomization [33].

Sampling Procedures

In the aforementioned parental randomized controlled trial, participants were recruited through multiple clinical centers via advertisements and clinician referrals [33]. Sampling followed consecutive enrollment of eligible patients [33]. The sample for this cost-effectiveness study comprises all randomized participants with complete outcome data in the parental trial [33].

Sample Size, Power, and Precision

The sample size was determined in the parent clinical trial based on detecting meaningful differences in insomnia severity (Insomnia Severity Index [ISI] scores) [33]. For the current economic evaluation, all available trial participants were included; therefore, no additional power analysis was required. Precision was improved through probabilistic sensitivity analyses using distributions for all parameters.

Measures and Covariates

Intervention and Comparator

The intervention is the digital CBT, Somzz, for insomnia treatment, and the comparator is CBT or pharmacotherapy without Somzz. As aforementioned, underpinning this study was a multicenter, single-blind, RCT (trial registration: KCT0007292) to assess the safety and efficacy of Somzz [33]. This study evaluates the cost-effectiveness of the Somzz using the collected and published information from the aforementioned parent clinical trial [33].

Outcomes

Outcomes included direct efficacy indicators related to insomnia treatment indicators (probabilities of remission, maintaining remission, attrition) and quality-adjusted-life-years. The direct efficacy values were extracted from the parent clinical trial [33]; (1) remission was defined as an ISI score of 7 or lower, indicating the absence of clinically significant insomnia; (2) the attrition; and (3) the probability of maintaining remission, which were measured during the follow-up at 18 weeks after device withdrawal. The minimum value of the attrition rate was the lower limit of the 95% CI, assuming that the attrition rate followed a normal distribution in the population. The maximum value was determined by applying values obtained from the literature [37], considering the possibility of an underestimated dropout rate in the clinical trial. QALYs were used as a health outcome measure in the analysis. QALYs were derived from the parent clinical trial [33], which measured QALY with the Short Form 6 Dimensions (SF-6D) Health State Short Form (a subset of 6 items from the Short Form Health Survey Version 1) [38]. The SF-6D includes 6 dimensions, each with 2 to 5 levels, and is capable of generating 7500 different health states. Brazier algorithm was used to calculate utility values [39]. The SF-6D was reported to show superior sensitivity to changes in mild-to-moderate physical and mental health conditions compared to the EQ-5D-3L [40,41].

Direct Medical Costs

We defined direct medical costs related to insomnia as expenses incurred from visits to clinics, hospitals, or pharmacies, specifically for insomnia treatment. The base case value for the treatment group was determined as the sum of 2 consultation fees, the cost of CBT software, and medication expenses. The minimum value for the treatment group was calculated as the 2 consultation fees, assuming no costs for CBT software or medication. The upper limit value for the treatment group was calculated by increasing the CBT software cost up to 200,000 KRW (US $155.28) with 2 consultations and related medications. The primary value for the control group included only medication costs, which were adjusted for adherence rates of 14% for prescription medications and 2% for over-the-counter medications. The minimum value of medication costs was calculated as the mean prescription medication cost subtracted from the SD. The maximum value was derived by adding the SD to the mean prescription medication cost. Medical expenses unrelated to insomnia treatment were classified as indirect medical costs.

Indirect Costs

Indirect costs were calculated as the sum of absenteeism, productivity loss, and workplace accident costs. Absenteeism and productivity losses caused by insomnia were monetized as part of absenteeism costs. We converted the number of days absent from work (cost of absenteeism) and the degree of productivity decline (cost of reduced work productivity) due to insomnia into a monetary cost for wage workers participating in the clinical trial [33]. Insomnia increases the likelihood of workplace accidents; therefore, the associated costs were estimated by multiplying the probability of insomnia-related accidents by the average cost per workplace accident. Insomnia can cause workplace accidents; therefore, we calculated the cost of insomnia-related accidents by multiplying the probability of an accident occurring because of insomnia by the average cost of an accident.

Data Collection

All clinical and quality-of-life data were collected prospectively during the parent RCT at 6 scheduled visits (screening, baseline, 1 week, 2 weeks, 4 weeks, and 18 weeks) [33]. Economic variables were derived from Korean national databases, trial-reported work impairment, and published literature. All data were de-identified before secondary analysis.

Quality of Measurements

Clinical outcomes (eg, ISI) were collected in the parent clinical trial by trained personnel following standardized protocols [33]. Economic parameters used validated national datasets. No missing data were reported in the parent clinical trial; no new missingness occurred in the secondary dataset.

Instrumentation

The instrumentation for this study included the Somzz dCBT-I mobile app, which delivers standardized CBT modules for insomnia through a smartphone-based digital platform, and the SF-6D survey instrument, a validated and widely used tool for measuring health-related utility values. All instruments were used in their original, validated formats, and no modifications were made to any components or measures.

Masking

The parent clinical trial used a single-blind design in which outcome assessors were blinded to treatment assignment. Participants were aware of their intervention due to the nature of the digital therapeutic. The economic evaluation was fully blinded to group identifiers [33].

Psychometrics

The SF-6D utilities were derived from SF-36 responses, and the psychometric properties of the Korean SF-36 support its use in Korean populations. In a Korean general population study (N=600), the SF-36 demonstrated acceptable internal consistency and test–retest reliability, with reliability coefficients ranging from 0.54 to 0.80 across domains [38,39,42]. In addition, the Korean Health Insurance Review and Assessment Service economic evaluation guideline allows the use of multi-attribute utility instruments, including SF-6D–derived utilities, in the Korean health technology assessment context [35]. Utility scoring used validated country-neutral algorithms. Attrition and remission measurements followed established ISI-based definitions [33].

Data Diagnostics

We evaluated plausibility ranges for all parameters, checked for outliers in cost distributions, and verified internal model consistency. The attrition rate distribution was modeled using normal assumptions based on trial data. Costs from claims databases were cross-validated with external Korean sources.

Analytic Strategy

Base-Case Analysis

The base-case analysis estimated the incremental cost-effectiveness ratio (ICER) by comparing the costs and health outcomes of digital CBT with those of treatment as usual. Health outcomes were measured in QALYs, and the analysis incorporated both direct medical costs and indirect societal costs to comprehensively assess the economic impact of the intervention.

Sensitivity Analysis

We ensured the robustness of our findings by varying several key assumptions, including the frequency of clinic visits, the need for additional consultations and educational sessions, and incremental health care expenditures related to insomnia. We also varied the financial impact of absenteeism, presenteeism, and workplace accidents resulting from insomnia. For absenteeism and presenteeism, we alternatively used data from the Work Productivity and Activity Impairment, which provides average, maximum, and minimum values for these factors [43]. We established the minimum cost of workplace accidents as zero with a scenario of no accident occurrence and the maximum cost as the highest individual accident cost recorded in the Industrial Accident Insurance Statistics [44] to account for the most severe financial impact observed.

Ethical Considerations

The trial was approved by the institutional review boards (IRBs) of all participating centers and conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines [33]. This study is a secondary analysis using existing data with primary consent, and the original consent or IRB approval covers secondary analysis without additional consent. Furthermore, ethical approval for the cost-effectiveness study was obtained from the Institutional Review Board of Yonsei University (approval number 202111-HR-2636-01). A waiver of informed consent for the use of retrospective data was also granted by the Institutional Review Board of Yonsei University. No financial incentives were provided in the parental clinical trial [33]. All data were deidentified before analysis. This study involved secondary analysis of anonymized clinical trial data already covered under the original IRB approval [33]. No images or supplementary materials in this article contain identifiable information about individual participants or users. All figures and tables present aggregated or schematic data only. Therefore, no additional consent for the use of identifiable images was required. If any materials containing identifiable information were to be included in the future, we would obtain explicit written consent from those individuals and upload the corresponding consent documentation in accordance with journal guidelines.

Results

Participant Flow

Overview

In the parental clinical trial, a total of 98 adults with chronic insomnia were enrolled and randomized in a 1:1 ratio to the Somzz group (n=49) or the control (sleep hygiene education [SHE]) group (n=49). All participants initiated the assigned intervention, and all participants had 6 scheduled visits (screening, baseline, 1 week, 2 weeks, 4 weeks, and 18 weeks as post intervention) over 6-9 months. The attrition rate in the Somzz group was 12% (6/49), whereas the SHE group showed comparable retention (5/49, 10%) over the 4-month follow-up period. Outcome analyses were conducted according to the intention-to-treat principle [33].

Recruitment

Participants were recruited from 3 university-affiliated hospitals through outpatient clinics and study advertisements. Eligible individuals were screened by research staff and invited to participate between baseline and randomization. After informed consent, participants completed baseline assessments and were randomly assigned to the Somzz or SHE condition using a single-blind allocation procedure [33].

Baseline Characteristics

At baseline, the 2 groups were comparable in demographic and clinical characteristics. Participants were adults with chronic insomnia and demonstrated similar insomnia severity, sleep diary metrics, and mental health profiles prior to randomization. No meaningful group differences were observed at baseline, supporting the validity of the randomization process [33].

Primary Outcomes

Outcomes included direct efficacy indicators related to insomnia treatment indicators (probabilities of remission, maintaining remission, and attrition) and quality-adjusted-life-years. The

direct efficacy values were extracted from the parent clinical trial [33]; (1) remission was 51% (22/43) of the digital CBT group and 14% (6/44) of the control group post intervention; (2) the attrition rate was 12% (6/49) for the in the digital CBT group and 10% (5/49) in the control group; (3) the probability of maintaining remission was 35% (15/43) of the treatment group and 11% (5/44) of the control group. For indirect medical costs, we compared medical claims from individuals diagnosed with insomnia to those without an insomnia diagnosis and determined that excess health expenditure caused by insomnia would amount to 553,625 KRW (US $429.83) over 9 weeks (Table 1).

Indirect costs were calculated as the sum of absenteeism, productivity loss, and workplace accident costs. In a parent clinical study, 54% were "individuals employed in paid work," and they reported being absent from work because of insomnia-related issues for 3.11 hours per week on average before starting treatment [33]. Therefore, we calculated the cost of absenteeism due to insomnia per one 9-week treatment cycle as follows: 9 weeks × 3.11h/week × 19,806 KRW (US $15.38) (hourly wage) × 54% (employment rate) ≒ 33,312 KRW (US $25.87). The parent clinical trial participants reported a 55% decline in productivity caused by insomnia [33]. Therefore, the indirect cost of lost productivity for one 9-week treatment cycle was approximated as follows: 45 days (5 days of work per week for the 9-week treatment period) × 55% (productivity loss due to insomnia) × 157,260 KRW (US $122.10)/day (daily wage) × 54% (employment rate) ≒ 2,101,347 KRW (US $1631.49) (Table 1). The probability of an insomnia-related workplace accident was assumed to be 1% [11]. According to the Ministry of Employment and Labor's Industrial Accident Insurance Statistics in Korea, the total amount of industrial accident insurance benefits in 2020 was 6.00 trillion KRW paid across 350,000 beneficiaries, resulting in an average of 17 million KRW per person [44]. Thus, the estimated accident cost per person is 85 million KRW annually and approximately 79,562 KRW (US $61.77) for a 9-week period [45,46] (Table 1).

Table 1. Model input parameters for the cost-effectiveness analysis of digital CBT (Somzz) for chronic insomnia among Korean adults.

Parameters[a]	Value assumed (range)	Distribution	Source
Outcomes			
Remission probability			
Treatment group (Somzz)	0.51 (0.40-0.68)	Beta	[33], ISI[b]<7; (minimum ISI < 6 - maximum ISI < 8)
Control group	0.14 (0.09-0.40)	Beta	[33]
Probability of maintaining remission for 18 weeks			
Treatment group (Somzz)	0.35 (0.18-0.47)	Beta	[33] (minimum ISI<6 - maximum ISI<8)
Control group	0.11 (0.09-0.18)	Beta	[33]
Attrition rate			
Treatment group (Somzz)	0.12 (0.08-0.30)	Beta	[33]; (minimum lower level of 95% CI- maximum [37])
Control group	0.10 (0.06-0.40)	Beta	[33]; (minimum [33] - maximum [31])
Quality-adjusted life years			
Insomnia	0.684 (0.460-1.0)	Beta	[33]
Noninsomnia	0.810 (0.573-1.0)	Beta	[33]
Costs			
Direct medical costs related to insomnia			
Treatment group (Somzz)	KRW 91,194 (29,100-237,194)[c]	Gamma	[47,48]
Control group	KRW 8,094 (7,890-10,068)[d]	Gamma	[47,48]
Direct medical costs unrelated to insomnia			
Treatment group (Somzz)	KRW 553,625 (0-136,074,635)[e]	gamma	[49]
Control group	KRW 437,689 (0-89,272,731)[f]	Gamma	[49]
Indirect costs			
Workplace absenteeism	KRW 33,312 (0-278,497)[g]	Gamma	[33], [45,46]
Workplace presenteeism	KRW 2,101,347 (0-3,827,195)[h]	Gamma	[33], [45,46]
Additional costs from workplace accidents	KRW 79,562 (0-1,590,541)[I]	Gamma	[33], [11,45,46]
Additional costs			
Digital CBT[j]	KRW 91,194 (29,100-237,194)[k]	Gamma	Internal data

[a]Model input parameters for the cost-effectiveness analysis of digital cognitive behavioral therapy (Somzz) for chronic insomnia among Korean adults were derived from clinical trial data and relevant literature. These parameters included remission and attrition probabilities, health-state utilities, direct and indirect costs, and the corresponding probability distributions used for probabilistic modeling.

[b]ISI: Insomnia Severity Index.

[c]US $70.80 (US $22.59-184.16).

[d]US $6.29 (US $6.13-7.82).

[e]US $429.83 (US $0-105648.35).

[f]US $339.82 (US $0-69311.35).

[g]US $25.86 (US $0-216.23).

[h]US $1631.49 (US $0-2971.43).

[i]US $61.77 (US $0-1234.90).

[j]CBT: cognitive behavioral therapy.

[k]US $70.80 (US $22.59-184.16).

Statistics and Data Analysis

Cost-Effectiveness of Digital CBT

Table 2 presents scenarios detailing the simulated health utility (QALYs) and cost outcomes for the baseline model. All scenarios, except Scenario 4, were analyzed by including only direct medical costs. Costs and QALYs increased for patients who implemented the digital CBT solution. The model predicted that the use of digital software to treat insomnia would produce 0.3701 QALYs and 101,729 KRW (US $78.98) of expected costs compared with 0.3609 QALYs and 22,038 KRW generated by treatment as usual without the aid of digital software over 6.5 months. Introducing the digital insomnia solution would increase costs by 79,691 KRW (US $61.87) and QALYs by 0.0091. The ICER associated with insomnia treatment using digital software was 8,719,727 KRW (US $6770.00) per QALY gained, which is well below the hypothetical threshold value in Korea of 30,000,000 KRW (US $23,292)/QALY gained.

Table 2. Cost-effectiveness analysis of digital cognitive behavioral therapy (Somzz) for chronic insomnia among Korean adults relative to the control group by base case and scenario.

Strategy	Cost, KRW (US $)	Effectiveness (QALYs)[a]	Incremental cost, KRW (US $)	Incremental effectiveness (QALYs)	ICER[b], KRW /QALYs (US $/QALYs)
Base case: 2 clinic visits					
Control group	22,038 (17.11)	0.3609	—[c]	—	—
Treatment group (Somzz)	101,729 (78.98)	0.3701	79,691 (61.88)	0.0091	8,719,727 (6,770.87)
Scenario 1: 3 clinic visits					
Control group	22,038 (17.11)	0.3609	—	—	—
Treatment group (Somzz)	113,859 (88.40)	0.3701	91,821 (71.30)	0.0091	10,046,980 (7,801.48)
Scenario 2: 2 clinic visits + educational and consultant fees					
Control group	22,038 (17.11)	0.3609	—	—	—
Treatment group (Somzz)	131,729 (102.27)	0.3701	109,691 (85.18)	0.0091	12,002,298 (9,319.78)
Scenario 3: 3 clinic visits + educational and consultant fees					
Control group	22,038 (17.11)	0.3609	—	—	—
Treatment group (Somzz)	143,859 (111.69)	0.3701	121,821 (94.59)	0.0091	13,329,551 (10,350.40)
Scenario 4: all costs included[d]					
Control group	7,679,657 (596.49)	0.3609	—	—	—
Treatment group (Somzz)	6,777,978 (526.42)	0.3701	–901,678 (–700.15)	0.0091	–98,660,780 (76,610.10)

[a]QALYs: quality-adjusted life years.

[b]ICER: incremental cost-effectiveness ratio.

[c]Not available.

[d]Direct medical costs related to insomnia + (educational and consultation fees) + direct medical costs unrelated to insomnia + indirect costs.

Compared to the control group, the ICER for insomnia treatment using the digital intervention, assuming all possible indirect costs, yielded a lower cost of 901,678 KRW (US $700.06) and higher QALYs of 0.0091. This negative ICER arises from combining the reduced total costs and increased health utility. Digital CBT reduces insomnia-related indirect costs, such as health care expenses and lost workplace productivity, through higher effectiveness and remission rates. These cost reductions surpass the direct costs associated with digital CBT (Table 2).

Sensitivity Analysis

We conducted a one-way deterministic and probabilistic sensitivity analysis to test the robustness of our findings. Figure 2 shows a tornado diagram of the results of this analysis. The direct costs of the digital CBT group, which ranged from 29,100 to 237,194 KRW (US $22.59 to US $184.16), had the largest impact on the ICER, followed by the probability of remission in the control group. These parameters had the greatest impact in all scenarios (Multimedia Appendices 2-4). The Monte Carlo simulation results indicated that digital CBT is the optimal cost-effective strategy with a probability of 67.4%, represented as an ICE scatterplot in Figure 3.

Figure 2. Tornado diagram for incremental cost-effectiveness ratio. ICER: incremental cost-effectiveness ratio.

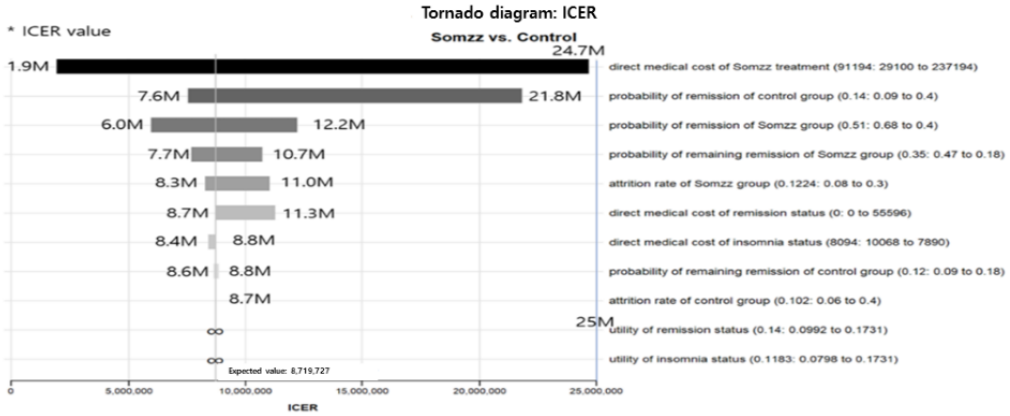
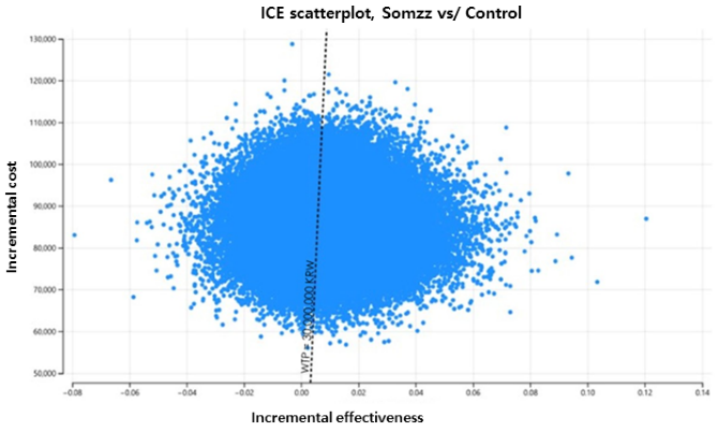


Figure 3. Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot (estimates based on the 95% CI).



For Scenarios 1-3, digital CBT was the most cost-effective strategy, with probabilities of 66.9%, 65.1%, and 63.9%, respectively (Multimedia Appendices 5-7). Furthermore, digital CBT was identified as the most cost-effective strategy at a willingness-to-pay threshold of 15,000,000 KRW (US $11,647.5), with probabilities of 58.8%, 57.8%, 53.5%, and 51.3% for the base case and Scenarios 1-3, respectively (Multimedia Appendix 8).

Discussion

Support of Original Hypotheses

We combined a decision tree and the Markov model to assess the cost-effectiveness of digital CBT for adults seeking treatment for insomnia in South Korea. While several prior studies have evaluated the cost-effectiveness of digital CBT for insomnia, most were conducted in Western health care systems [29-32]. In contrast, this study evaluates a regulated, commercially available digital therapeutic using Korean randomized trial data and explicitly incorporates societal costs that are highly relevant to the Korean labor and health care context. From a health care perspective, digital CBT slightly increased health care costs while improving health outcomes compared with standard care. In contrast, from a societal perspective, digital CBT was cost-beneficial, yielding a negative ICER. Accordingly, the digital therapeutic evaluated in this study was highly cost-effective for functional outcomes in insomnia treatment at a willingness-to-pay threshold of 30,000,000 KRW (US $23,295) per QALY gained and remained cost-effective at a lower threshold of 15,000,000 KRW (US $11,647.5) per QALY gained. These findings underscore that incorporating broader societal costs meaningfully changes the economic value of insomnia treatment interventions. Reduced health care use, lower risks of workplace accidents, and improved work productivity associated with insomnia remission were key drivers of cost-effectiveness. Both deterministic and probabilistic sensitivity analyses confirmed the robustness of these results.

Similarity of Results and Interpretation

These results align with those of previous studies that found guided digital CBT to be cost-effective [30-32]. One strength of our study was the inclusion of indirect costs. We considered reductions in health care expenditure as effective insomnia treatment can lead to fewer medical visits, less need for medication, and lower overall health care use [21,22]. Furthermore, we accounted for a decreased risk of workplace accidents. Insomnia impairs cognitive and physical functioning, which increases the likelihood of errors and accidents at work [50,51]. Treating insomnia can mitigate this risk, thereby leading to safer work environments. Our analysis also included improved workplace productivity. Insomnia negatively affects concentration, efficiency, and overall job performance [52,53]. Therefore, effective treatment can enhance productivity by improving employees' sleep quality and daytime functioning. By capturing these indirect costs, our analysis demonstrates the broader economic benefits of insomnia treatment beyond

immediate improvements in individual health. Insomnia is linked to 7.2% of all costly workplace accidents and errors, and accounts for 23.7% of the total costs associated with these incidents [11]. Not including these additional costs could underestimate the total economic burden of insomnia and potential costs saved through the application of effective treatment. In contrast, including costs related to workplace accidents and errors could reveal higher overall costs, and thus, a greater potential for cost savings through insomnia treatment.

The demand for effective insomnia treatment is substantial, with 20%-30% of the population experiencing insomnia symptoms [1-3]. However, several factors hinder traditional clinician-delivered CBT, including the availability of trained therapists and the need for patients to attend scheduled sessions, which many people can find challenging. Patients often face difficulties in taking time off work and securing childcare or older adult care to attend in-person therapy sessions. These logistical challenges limit the reach and effectiveness of clinician-delivered CBT [54].

Digital CBT delivery offers a promising solution to these accessibility issues by enabling instant and unlimited access to CBT resources, thereby allowing patients to engage with therapy at any time and from any location, effectively eliminating waiting times. This flexibility is crucial for patients who may be unable to attend regular appointments because of their busy schedules [54]. Digital platforms can make CBT more widely available, ensuring that more individuals who require treatment can access it promptly and conveniently. This approach not only expands the reach of effective insomnia treatment but also aligns with the high demand for such interventions, ultimately helping bridge the gap between the recommended use of CBT and its actual accessibility and use.

Generalizability

The generalizability of this study's findings should be interpreted in the context of the Korean population and health care system. Participants were digitally literate adults recruited from university hospitals [33], which may not fully represent individuals with limited technology access or older adults who may engage differently with mobile health interventions. Nonetheless, given Korea's high smartphone penetration rate and increasing familiarity with digital health tools, Somzz has strong potential for scalable implementation across the broader adult population.

Economic estimates in this model reflect Korean wage levels, health care use patterns, and societal cost structures. Although absolute cost differences may vary in other countries, the primary mechanisms driving cost-effectiveness, such as higher remission rates, reduced health care use, and improved productivity, are likely applicable to other high-income settings with similar burdens of insomnia and labor market characteristics. Because Somzz was developed within the Korean cultural and linguistic environment, differences in cultural attitudes toward sleep, mental health, and digital therapeutics may influence usability and adherence in other populations. Additionally, real-world engagement with digital CBT-I may vary outside a controlled trial setting, which could influence long-term effectiveness and economic outcomes.

Overall, while these findings are most directly applicable to technologically advanced health care systems with strong digital readiness, they provide valuable insights for other countries considering digital CBT-I adoption. Extrapolation should be undertaken cautiously with attention to cultural, economic, and infrastructural differences.

Implications

This study provides novel evidence by incorporating indirect societal costs, including productivity loss and workplace accidents, into the economic evaluation of digital CBT, dimensions that are rarely captured in prior analyses. Importantly, by focusing on a digital therapeutic that meets regulatory standards and is designed for real-world implementation, this study extends prior digital CBT economic evaluations toward policy-relevant evidence for reimbursable digital therapeutics. By using Korean clinical trial data [33] and locally relevant economic parameters rather than relying on Western-based assumptions, the study offers population-specific insights that more accurately reflect real-world conditions. This comprehensive approach strengthens the relevance of our findings for clinicians, policymakers, and employers by demonstrating the broader economic and workplace impact of digital CBT and informing decision-making for large-scale adoption within the Korean health care and occupational health systems.

One limitation of this study was the relatively short follow-up period. Although a period of 6.5 months is approximately the same or longer than that used in all previous studies on the cost-effectiveness of CBT for insomnia, it may not be sufficient to fully capture the long-term economic benefits of treatment. Simulating costs over a longer period, such as 12 or 24 months, could more effectively demonstrate potential long-term cost reductions for employers. A longer time horizon may allow for a more significant accumulation of treatment benefits, suggesting that our current results may be conservative.

Another limitation is the sample size. Economic evaluations conducted in parallel with clinical trials often face power issues because cost variables typically have a higher variance and require larger sample sizes than clinical outcomes to achieve statistical significance. Considering these challenges, we adopted a probabilistic decision-making approach to account for the uncertainty and variability in cost and utility estimates. This approach helped us perform a more robust analysis despite the limitations of the sample size and follow-up duration in this study.

Finally, the parent clinical trial from which outcome data were extracted [33] was not structured to compare traditional and digital CBT directly, primarily because of practical constraints (eg, despite having insomnia symptoms, only a small proportion of patients actively seek pharmacotherapy in the real world). This approach mimics real-world scenarios in which patients often receive a combination of treatments rather than a single isolated therapy.

This study did not involve direct patient or stakeholder engagement. However, the societal burden and impact of insomnia were incorporated into the analytical framework. No

direct involvement from patients, service recipients, the general public, or other stakeholders influenced the approach or findings of this study.

Conclusions

The results of this study suggest the significant potential of early integration of digital CBT into insomnia treatment to improve health care outcomes and generate substantial cost-effectiveness. These findings underscore the importance of incorporating such technology as a valuable enhancement to routine medical practice, thereby ultimately enhancing overall insomnia care. Beyond demonstrating cost-effectiveness, this study contributes to the growing evidence base supporting digital therapeutics by illustrating how locally grounded economic evaluations can inform reimbursement, employer adoption, and national digital health policy. Given these promising results, future research should focus on expanding access to digital CBT and assessing its effectiveness on a population-wide scale. Understanding the effectiveness of digital CBT in real-world settings, in which factors such as varying levels of access, patient engagement, and implementation challenges can influence outcomes, is crucial.

Acknowledgments

We did not use generative artificial intelligence to write any portion of the manuscript.

Funding

This work was supported by the Ministry of Health and Welfare (RS-2023-00266398) and the Hankuk University of Foreign Studies Research Fund. The funder had no involvement in the study design, data analysis, interpretation of the results, writing of this article, or the decision to submit it for publication.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request. Most of the input data used to derive the results of this study are extracted from published literature, which is all listed in the Reference section with corresponding citations in the text.

Conflicts of Interest

Jimin Woo is affiliated with AIMNEXT Co., Ltd.

Multimedia Appendix 1
Supplementary Table 1.
[PDF File (Adobe PDF File), 65 KB - mhealth_v14i1e71750_app1.pdf]

Multimedia Appendix 2
Deterministic sensitivity analysis results: Scenario 1 – tornado diagram for ICER.
[PNG File , 315 KB - mhealth_v14i1e71750_app2.png]

Multimedia Appendix 3
Deterministic sensitivity analysis results: Scenario 2 – tornado diagram for ICER.
[PNG File , 303 KB - mhealth_v14i1e71750_app3.png]

Multimedia Appendix 4
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 3.
[PNG File , 317 KB - mhealth_v14i1e71750_app4.png]

Multimedia Appendix 5
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 1.
[PNG File , 369 KB - mhealth_v14i1e71750_app5.png]

Multimedia Appendix 6
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 2.
[PNG File , 338 KB - mhealth_v14i1e71750_app6.png]

Multimedia Appendix 7
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot for Scenario 3.

[PNG File , 361 KB - mhealth_v14i1e71750_app7.png]

Multimedia Appendix 8
Probabilistic sensitivity analysis results: incremental cost-effectiveness scatterplot: (A) base case; (B) Scenario 1; (C) Scenario 2; and (D) Scenario 3 (WTP = 15,000,000).
[PNG File , 528 KB - mhealth_v14i1e71750_app8.png]

References

1. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest 2014;146(5):1387-1394. [doi: 10.1378/chest.14-0970] [Medline: 25367475]
2. Qaseem A, Kansagara D, Forciea MA, Cooke M, Denberg TD, Clinical Guidelines Committee of the American College of Physicians. Management of chronic insomnia disorder in adults: a clinical practice guideline from the American College of Physicians. Ann Intern Med 2016;165(2):125-133 [FREE Full text] [doi: 10.7326/M15-2175] [Medline: 27136449]
3. Perlis ML, Posner D, Riemann D, Bastien CH, Teel J, Thase M. Insomnia. Lancet 2022;400(10357):1047-1060. [doi: 10.1016/S0140-6736(22)00879-0] [Medline: 36115372]
4. Morin CM, Bjorvatn B, Chung F, Holzinger B, Partinen M, Penzel T, et al. Insomnia, anxiety, and depression during the COVID-19 pandemic: an international collaborative study. Sleep Med 2021;87:38-45 [FREE Full text] [doi: 10.1016/j.sleep.2021.07.035] [Medline: 34508986]
5. Zhang C, Yang L, Liu S, Ma S, Wang Y, Cai Z, et al. Survey of insomnia and related social psychological factors among medical staff involved in the 2019 novel coronavirus disease outbreak. Front Psychiatry 2020;11:306 [FREE Full text] [doi: 10.3389/fpsyt.2020.00306] [Medline: 32346373]
6. Wade AG. The societal costs of insomnia. Neuropsychiatr Dis Treat 2010;7:1-18 [FREE Full text] [doi: 10.2147/NDT.S15123] [Medline: 21326650]
7. Staner L. Comorbidity of insomnia and depression. Sleep Med Rev 2010 Mar;14(1):35-46. [doi: 10.1016/j.smrv.2009.09.003] [Medline: 19939713]
8. Ge L, Guyatt G, Tian J, Pan B, Chang Y, Chen Y, et al. Insomnia and risk of mortality from all-cause, cardiovascular disease, and cancer: systematic review and meta-analysis of prospective cohort studies. Sleep Med Rev 2019;48:101215. [doi: 10.1016/j.smrv.2019.101215] [Medline: 31630016]
9. Bolge SC, Doan JF, Kannan H, Baran RW. Association of insomnia with quality of life, work productivity, and activity impairment. Qual Life Res 2009;18(4):415-422. [doi: 10.1007/s11136-009-9462-6] [Medline: 19288223]
10. Garbarino S, Magnavita N, Guglielmi O, Maestri M, Dini G, Bersi FM, et al. Insomnia is associated with road accidents. Further evidence from a study on truck drivers. PLoS One 2017;12(10):e0187256 [FREE Full text] [doi: 10.1371/journal.pone.0187256] [Medline: 29088276]
11. Shahly V, Berglund PA, Coulouvrat C, Fitzgerald T, Hajak G, Roth T, et al. The associations of insomnia with costly workplace accidents and errors: results from the America insomnia survey. Arch Gen Psychiatry 2012;69(10):1054-1063. [doi: 10.1001/archgenpsychiatry.2011.2188] [Medline: 23026955]
12. Riemann D, Baglioni C, Bassetti C, Bjorvatn B, Dolenc Groselj L, Ellis JG, et al. European guideline for the diagnosis and treatment of insomnia. J Sleep Res 2017;26(6):675-700. [doi: 10.1111/jsr.12594] [Medline: 28875581]
13. Wilson S, Anderson K, Baldwin D, Dijk D, Espie A, Espie C, et al. British Association for Psychopharmacology consensus statement on evidence-based treatment of insomnia, parasomnias and circadian rhythm disorders: an update. J Psychopharmacol 2019;33(8):923-947. [doi: 10.1177/0269881119855343] [Medline: 31271339]
14. Sateia MJ, Buysse DJ, Krystal AD, Neubauer DN, Heald JL. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline. J Clin Sleep Med 2017;13(2):307-349 [FREE Full text] [doi: 10.5664/jcsm.6470] [Medline: 27998379]
15. Glass J, Lanctôt KL, Herrmann N, Sproule BA, Busto UE. Sedative hypnotics in older people with insomnia: meta-analysis of risks and benefits. BMJ 2005 Nov;331(7526):1169 [FREE Full text] [doi: 10.1136/bmj.38623.768588.47] [Medline: 16284208]
16. Tannenbaum C, Diaby V, Singh D, Perreault S, Luc M, Vasiliadis H. Sedative-hypnotic medicines and falls in community-dwelling older adults: a cost-effectiveness (decision-tree) analysis from a US Medicare perspective. Drugs Aging 2015;32(4):305-314. [doi: 10.1007/s40266-015-0251-3] [Medline: 25825121]
17. Choi Y, Lee MH, Choi J, Kim S, Kim J, Lee YJ. Survey of insomnia treatment status for doctors. Sleep Med Psychophysiol 2016;23(2):77-83. [doi: 10.14401/kasmed.2016.23.2.77]
18. Kaufmann CN, Spira AP, Alexander GC, Rutkow L, Mojtabai R. Trends in prescribing of sedative-hypnotic medications in the USA: 1993-2010. Pharmacoepidemiol Drug Saf 2016;25(6):637-645 [FREE Full text] [doi: 10.1002/pds.3951] [Medline: 26711081]
19. Miller CB, Valenti L, Harrison CM, Bartlett DJ, Glozier N, Cross NE, et al. Time trends in the family physician management of insomnia: the Australian experience (2000-2015). J Clin Sleep Med 2017;13(6):785-790 [FREE Full text] [doi: 10.5664/jcsm.6616] [Medline: 28454597]

20. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. JAMA Psychiatry 2017;74(1):68-75. [doi: 10.1001/jamapsychiatry.2016.3249] [Medline: 27902836]
21. Daley M, Morin C, LeBlanc M, Grégoire J, Savard J. The economic burden of insomnia: direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers. Sleep 2009;32(1):55-64. [doi: 10.5665/sleep/32.1.55]
22. Kessler R, Berglund P, Coulouvrat C, Hajak G, Roth T, Shahly V, et al. Insomnia and the performance of US workers: results from the America insomnia survey. Sleep 2011;34(9):1161-1171 [FREE Full text] [doi: 10.5665/SLEEP.1230] [Medline: 21886353]
23. Sweetman A, Reynolds C, Richardson C. Digital cognitive behavioural therapy for insomnia versus digital sleep education control in an Australian community-based sample: a randomised controlled trial. Intern Med J 2024;54(11):1838-1848. [doi: 10.1111/imj.16521] [Medline: 39257295]
24. Seyffert M, Lagisetty P, Landgraf J, Chopra V, Pfeiffer PN, Conte ML, et al. Internet-delivered cognitive behavioral therapy to treat insomnia: a systematic review and meta-analysis. PLoS One 2016;11(2):e0149139 [FREE Full text] [doi: 10.1371/journal.pone.0149139] [Medline: 26867139]
25. Ye Y, Chen N, Chen J, Liu J, Lin L, Liu Y, et al. Internet-based cognitive-behavioural therapy for insomnia (ICBT-i): a meta-analysis of randomised controlled trials. BMJ Open 2016;6(11):e010707 [FREE Full text] [doi: 10.1136/bmjopen-2015-010707] [Medline: 27903557]
26. Zachariae R, Lyby MS, Ritterband LM, O'Toole MS. Efficacy of internet-delivered cognitive-behavioral therapy for insomnia - a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2016;30:1-10. [doi: 10.1016/j.smrv.2015.10.004] [Medline: 26615572]
27. Lin W, Li N, Yang L, Zhang Y. The efficacy of digital cognitive behavioral therapy for insomnia and depression: a systematic review and meta-analysis of randomized controlled trials. PeerJ 2023;11:e16137 [FREE Full text] [doi: 10.7717/peerj.16137] [Medline: 37927792]
28. Zettor D, Endomba FT, Pierandrei A, Pinoit J, Chauvet-Gelinier J, Forestier N, et al. Effectiveness of digital cognitive behavioral therapy for insomnia on professional activity: a systematic review and meta-analysis of randomized controlled trials. Sleep Med Rev 2025;79:102024. [doi: 10.1016/j.smrv.2024.102024] [Medline: 39571405]
29. Fang L, Lyu Z, Ai S, Du S, Zhou W, Zeng S, et al. Is cognitive behavioral therapy for insomnia more cost-effective? New-perspective on economic evaluations: a systematic review and meta-analysis. Sleep 2024;47(8). [doi: 10.1093/sleep/zsae122] [Medline: 38795362]
30. Buntrock C, Lehr D, Smit F, Horvath H, Berking M, Spiegelhalder K, et al. Guided internet-based cognitive behavioral therapy for insomnia: health-economic evaluation from the societal and public health care perspective alongside a randomized controlled trial. J Med Internet Res 2021;23(5):e25609 [FREE Full text] [doi: 10.2196/25609] [Medline: 34028361]
31. Darden M, Espie CA, Carl JR, Henry AL, Kanady JC, Krystal AD, et al. Cost-effectiveness of digital cognitive behavioral therapy (Sleepio) for insomnia: a Markov simulation model in the United States. Sleep 2021;44(4). [doi: 10.1093/sleep/zsaa223] [Medline: 33151330]
32. Thiart H, Ebert D, Lehr D, Nobis S, Buntrock C, Berking M, et al. Internet-based cognitive behavioral therapy for insomnia: a health economic evaluation. Sleep 2016;39(10):1769-1778 [FREE Full text] [doi: 10.5665/sleep.6152] [Medline: 27450686]
33. Shin J, Kim S, Lee J, Gu H, Ahn J, Park C, et al. Efficacy of mobile app-based cognitive behavioral therapy for insomnia: multicenter, single-blind randomized clinical trial. J Med Internet Res 2024;26:e50555 [FREE Full text] [doi: 10.2196/50555] [Medline: 39058549]
34. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. Med Decis Making 1993;13(4):322-338. [doi: 10.1177/0272989X9301300409] [Medline: 8246705]
35. Guidelines for pharmaceutical economic evaluation [Web page in Korean]. HIRA OAK Repository. 2021. URL: https://repository.hira.or.kr/handle/2019.oak/2541 [accessed 2022-10-10]
36. Husereau D, Drummond M, Augustovski F, de Bekker-Grob E, Briggs AH, Carswell C, et al. Consolidated health economic evaluation reporting standards 2022 (CHEERS 2022) statement: updated reporting guidance for health economic evaluations. Value Health 2022;25(1):3-9 [FREE Full text] [doi: 10.1016/j.jval.2021.11.1351] [Medline: 35031096]
37. Karlin BE, Trockel M, Taylor CB, Gimeno J, Manber R. National dissemination of cognitive behavioral therapy for insomnia in veterans: therapist- and patient-level outcomes. J Consult Clin Psychol 2013;81(5):912-917. [doi: 10.1037/a0032554] [Medline: 23586730]
38. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. Med Care 2004;42(9):851-859. [doi: 10.1097/01.mlr.0000135827.18610.0d] [Medline: 15319610]
39. Mukuria C, Rowen D, Mulhern B, McDool E, Kharroubi S, Bjorner JB, et al. The short form 6 dimensions (SF-6D): development and evolution. Appl Health Econ Health Policy 2025;23(1):19-33 [FREE Full text] [doi: 10.1007/s40258-024-00919-8] [Medline: 39460886]
40. Brazier J, Roberts J, Tsuchiya A, Busschbach J. A comparison of the EQ-5D and SF-6D across seven patient groups. Health Econ 2004;13(9):873-884 [FREE Full text] [doi: 10.1002/hec.866] [Medline: 15362179]

41. Brazier J, Connell J, Papaioannou D, Mukuria C, Mulhern B, Peasgood T, et al. A systematic review, psychometric analysis and qualitative assessment of generic preference-based measures of health in mental health populations and the estimation of mapping functions from widely used specific measures. Health Technol Assess 2014;18(34):vii-viii, xiii [FREE Full text] [doi: 10.3310/hta18340] [Medline: 24857402]
42. Kim SH, Jo M, Lee S. Psychometric properties of the Korean short form-36 health survey version 2 for assessing the general population. Asian Nurs Res (Korean Soc Nurs Sci) 2013;7(2):61-66 [FREE Full text] [doi: 10.1016/j.anr.2013.03.001] [Medline: 25029923]
43. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. PharmacoEconomics 1993;4(5):353-365. [doi: 10.2165/00019053-199304050-00006]
44. Industrial Accident Insurance Annual Report. Industrial Accident Insurance Annual Report. 2023. URL: https://www.moel.go.kr/info/publict/publictDataView.do?bbs_seq=20241001451 [accessed 2022-10-10]
45. KOSIS. Insurance Benefit Payment Status (By Year and Business Size). Insurance Benefit Payment Status (By Year and Business Size). URL: https://tinyurl.com/mphycv28 [accessed 2022-10-10]
46. Park C, Lee H. A study on the economic loss costs related to industrial accidents - focusing on the manufacturing industry [Web page in Korean]]. Korea Labor Institute Research Report. 2018. URL: https://www.kli.re.kr/board.es?mid=a10505020000&bid=0007&act=view&list_no=138920&tag=&nPage=12 [accessed 2022-10-10]
47. 2022 Health institution fee schedule and patient co-payment table. HIRA.: Health Insurance Review & Assessment Service URL: https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020002000100&brdScnBltNo=4&brdBltNo=9293&pageIndex=1#none [accessed 2022-10-10]
48. Current list of reimbursed drugs and the reimbursement ceiling amount table. Health Insurance Review & Assessment Service. 2022. URL: https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA030014050000 [accessed 2022-10-10]
49. Sarsour K, Kalsekar A, Swindle R, Foley K, Walsh JK. The association between insomnia severity and healthcare and productivity costs in a health plan sample. Sleep 2011;34(4):443-450 [FREE Full text] [doi: 10.1093/sleep/34.4.443] [Medline: 21461322]
50. Uehli K, Mehta AJ, Miedinger D, Hug K, Schindler C, Holsboer-Trachsler E, et al. Sleep problems and work injuries: a systematic review and meta-analysis. Sleep Med Rev 2014;18(1):61-73 [FREE Full text] [doi: 10.1016/j.smrv.2013.01.004] [Medline: 23702220]
51. Wickwire EM, Shaya FT, Scharf SM. Health economics of insomnia treatments: the return on investment for a good night's sleep. Sleep Med Rev 2016;30:72-82. [doi: 10.1016/j.smrv.2015.11.004] [Medline: 26874067]
52. Hafner M, Stepanek M, Taylor J, Troxel W, van Stolk C. Why sleep matters-the economic costs of insufficient sleep: a cross-country comparative analysis. Rand Health Q 2017;6(4):11 [FREE Full text] [Medline: 28983434]
53. Rosekind M, Gregory K, Mallis M, Brandt S, Seal B, Lerner D. The cost of poor sleep: workplace productivity loss and associated costs. J Occup Environ Med 2010;52(1):91-98. [doi: 10.1097/JOM.0b013e3181c78c30] [Medline: 20042880]
54. Koffel E, Bramoweth AD, Ulmer CS. Increasing access to and utilization of cognitive behavioral therapy for insomnia (CBT-I): a narrative review. J Gen Intern Med 2018;33(6):955-962 [FREE Full text] [doi: 10.1007/s11606-018-4390-1] [Medline: 29619651]

Abbreviations

CBT: cognitive behavioral therapy
CHEERS: Consolidated Health Economic Evaluation Reporting Standards
dCBT-I: digital cognitive behavioral therapy for insomnia
ICER: incremental cost-effectiveness ratio
IRB: institutional review board
ISI: Insomnia Severity Index
QALY: quality-adjusted life years
RCT: randomized controlled trial
SF-6D: Short Form 6 Dimensions

Edited by S Brini; submitted 25.Jan.2025; peer-reviewed by A Baharav, D Xiao; comments to author 10.Jun.2025; revised version received 21.Dec.2025; accepted 25.Dec.2025; published 19.Jan.2026.

Please cite as:
Kim JH, Sohn M, Woo J, Shin J, Han E
Digital Cognitive Behavioral Therapy for Chronic Insomnia in South Korea: Cost-Effectiveness Analysis Using Decision Tree and Markov Modeling Based on a Secondary Analysis of a Randomized Clinical Trial
JMIR Mhealth Uhealth 2026;14:e71750
URL: https://mhealth.jmir.org/2026/1/e71750
doi:10.2196/71750
PMID:

Original Paper

Efficacy and Safety of Mobile App–Based Metamemory Cognitive Training for Mild Cognitive Impairment: Multicenter Randomized Clinical Trial

Sunyoung Kang[1], MD; Jung-In Lim[2], MA; Lukas Stenzel[3], PhD; Keun You Kim[1], MD, PhD; Eosu Kim[1,4,5], MD, PhD; Hong Jun Jeon[6], MD; Doo-Heum Park[6], MD, PhD; Hyun Kook Lim[7,8], MD, PhD; YongSoo Shim[9], MD, PhD; Jae-Won Jang[10], MD, PhD; Yeshin Kim[10], MD, PhD; Sanghoon Lee[11], BS; Kee Hyung Park[12], MD, PhD

[1]Department of Psychiatry & Institute of Behavioral Science in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Republic of Korea

[3]Cogthera GmbH, Munich, Germany

[4]Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Republic of Korea

[5]Metabolism-Dementia Research Institute, Yonsei University College of Medicine, Seoul, Republic of Korea

[6]Department of Psychiatry, Konkuk University Medical Center, Seoul, Republic of Korea

[7]Department of Psychiatry, Yeouido St. Mary's Hospital, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea

[8]Catholic Medical Center Institute for Basic Medical Science, The Catholic University of Korea Catholic Medical Center, Seoul, Republic of Korea

[9]Department of Neurology, St. Vincent's Hospital, College of Medicine, Catholic University of Korea, Seoul, Republic of Korea

[10]Department of Neurology, Kangwon National University Hospital, College of Medicine, Kangwon National University, Chuncheon, Gangwon-do, Republic of Korea

[11]Emocog Inc., Seoul, Republic of Korea

[12]Department of Neurology, College of Medicine, Gachon University Gil Medical Center, Incheon, Republic of Korea

Corresponding Author:
Kee Hyung Park, MD, PhD
Department of Neurology, College of Medicine
Gachon University Gil Medical Center
21 Namdong-daero 774beon-gil, Namdong-gu
Incheon, 21565
Republic of Korea
Phone: 82 32 460 8683
Email: khpark@gachon.ac.kr

Abstract

Background: Metamemory training (MMT) offers a potential nonpharmacological approach to enhance cognitive function in individuals with mild cognitive impairment (MCI). While digital cognitive training improves accessibility, the effectiveness of mobile app–based MMT has not been evaluated in a randomized clinical trial.

Objective: We aimed to evaluate the efficacy and safety of a mobile app–based MMT program, ET-101 (Cogthera), compared to a sham device control group in individuals with MCI.

Methods: This multicenter, randomized controlled trial enrolled participants with MCI, recruited from 7 medical centers, and randomly assigned them to the ET-101 or control group (1:1 ratio). The intervention lasted 12 weeks, with a 12-week follow-up. The ET-101 group received metamemory-based multimemory strategy training and real-time feedback. Assessments of cognition, the daily activities of living, and the quality of life were conducted at baseline, week 12, and week 24. The primary outcome was the proportion of participants who showed cognitive improvement as assessed by the Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog)-14 at weeks 12 and 24. Secondary outcomes included changes in the scores of scales assessing cognition, daily activities, and quality of life. Safety analysis assessed adverse events and their relation to digital therapeutics.

Results: In the full analysis set, 49 participants were included in the ET-101 group and 50 in the control group. At week 24, the proportion of responders who maintained or improved their ADAS-Cog-14 scores was significantly higher in the ET-101 group than in the control group (P=.002). Additionally, the ET-101 group showed a significant improvement in ADAS-Cog-14

scores at week 24 compared to baseline levels (estimates=–2.53; t_{265}=–3.05; Bonferroni-adjusted *P*=.003). A subdomain analysis revealed significant improvements in the memory (estimates=–2.50; t_{264}=–4.03; Bonferroni-adjusted *P*<.001) and language (estimates=–0.807; t_{290}=–3.68; Bonferroni-adjusted *P*<.001) domains at week 24 in the ET-101 group compared to the control group. In the safety analysis, 6 adverse events occurred in the ET-101 group and 4 in the control group, but none were related to the interventions. The attrition rate in the ET-101 group was 22.4% (11/49).

Conclusions: ET-101 significantly improved cognitive function compared to the sham device, with effects observed not only in the memory domain but also in the language domain, indicating a transfer effect. Therefore, ET-101 has the potential to provide effective MMT to a broader population with MCI by overcoming location and personnel limitations through a mobile app–based platform.

Trial Registration: ClinicalTrials.gov NCT05938426; https://clinicaltrials.gov/study/NCT05938426

(JMIR Mhealth Uhealth 2026;14:e73464) doi:10.2196/73464

KEYWORDS

cognitive decline; cognitive training; digital therapeutics; mild cognitive impairment; mobile app–based cognitive training

Introduction

Background

With the global population aging rapidly, the number of people affected by dementia is estimated to reach 153 million by 2050 [1]. Increasing evidence shows that timely diagnosis and treatment can slow the rate of cognitive deterioration, significantly reducing the burden on caregivers and social care systems [1]. As a result, effective interventions for mild cognitive impairment (MCI), which is at high risk of progressing, have become a priority in geriatric care [2].

Notably, anti–amyloid monoclonal antibody treatments, which have recently attracted attention in the management of MCI, may be appropriate only for patients who meet several eligibility criteria. Moreover, this treatment is associated with adverse effects that could limit continued use in some cases [3]. In contrast, cognitive training represents a relatively safe intervention with greater generalizability and applicability in the treatment of MCI, and it can also be used in combination with pharmacotherapy or as a stand-alone treatment option for patients who are unable to undergo or continue pharmacological therapy. According to a meta-analysis, cognitive training significantly improves memory function [4]. Moreover, magnetic resonance imaging–based studies have demonstrated that cognitive training improves cerebral blood flow and brain network connectivity, confirming the biological effects of cognitive interventions [5]. However, while cognitive training, as used in previous studies, improved older adults' performance on practiced tasks, the studies did not consider the participants' ability to self-perceive and regulate cognitive strategies during cognitively demanding tasks [6]. In addition, previous studies indicate that one of the key challenges in cognitive training for MCI is the lack of sustained effects once the training period ends [7,8]. These studies address the need for future interventions that aim to address and overcome this limitation and thereby enhance the real-life applicability of the learned strategies [7,8].

To address the limitations of previous methods of cognitive training, Flavell's concept of "metamemory" can be beneficial [9]. Metamemory refers to one's awareness of one's own memory, including the contents and processes of their memory system [6,9]. Metamemory training (MMT) takes into account the concepts of meta-knowledge, meta-monitoring, and meta-judgement [10]. Meta-knowledge addresses how aging affects memory abilities, enabling older adults to learn strategies to deal with age-related cognitive decline [10]. In the meta-monitoring and meta-judgment components, older adults assess their own performance, enabling them to regulate their memorization ability by independently adjusting effective memory strategies according to the demands of the task at hand [10]. According to previous studies, MMT not only shows positive effects on everyday memory performance in older populations but also improves global cognitive functions while significantly reducing the frequency of memory-related discomfort in daily life [11-14].

Conventionally, the main limitation of cognitive interventions for MCI is their reliance on face-to-face interactions with professionals, which hampers the accessibility and continuity of training [15]. However, digitalization is ideal for facilitating the migration of care from clinics to patients' homes, thereby overcoming constraints related to time and space. According to previous studies that evaluated the feasibility of digital cognitive therapy for MCI, patients adhered more to digital therapy compared to paper-and-pencil training. Additionally, in terms of effectiveness, a review of studies on computerized cognitive training for MCI found that its effect on cognitive function was not inferior to that of conventional cognitive interventions [15].

Objectives

Thus, metamemory-based digital cognitive training can be expected to enhance the benefits of cognitive training in individuals with MCI by addressing the limitations of previous cognitive interventions. Based on this concept, ET-101 (Cogthera)—a mobile app–based MMT program—was developed. In this randomized clinical trial, we aimed to evaluate the efficacy and safety of ET-101 in older individuals with MCI compared to a control group using a sham device. To evaluate both the immediate and sustained effects of ET-101, we assessed its efficacy compared to sham controls at 2 time points that is, at 12 weeks, immediately after completion of the intervention, and 24 weeks, corresponding to a 12-week follow-up after the end of the intervention. This study aimed to evaluate the

following hypotheses: first, for the primary outcome, we compared responder proportions (defined as participants showing maintenance or improvement in Alzheimer's Disease Assessment Scale-cognitive Cognitive Subscale [ADAS-Cog]-14 scores) between the ET-101 and sham control groups, hypothesizing that the responder rate would be higher in the ET-101 group. Second, regarding secondary outcomes, we assessed changes from baseline to 12- and 24-week follow-ups in validated measures of cognitive function, activities of daily living (ADL), and quality of life, hypothesizing that greater improvements would be observed in the ET-101 group versus the sham control group. Finally, exploratory analyses were conducted to determine whether improvements on ADAS-Cog-14 were predominantly observed in the memory, language, or praxis domains to identify which specific cognitive functions were most responsive to ET-101.

Methods

Study Design and Participants

This was a multicenter, 24-week parallel, randomized, sham device–controlled study of ET-101. The participants were recruited through memory clinics of 7 medical centers in the Republic of Korea. The participants were recruited from June to December 2023. Participants who met the following inclusion criteria were included in the study: (1) individuals aged 55-85 years; (2) patients diagnosed with amnestic MCI by trained psychiatrists or neurologists according to Petersen's criteria, characterized by cognitive complaints and objective memory impairment determined by either the Consortium to Establish a Registry for Alzheimer's Disease Neuropsychological Assessment Battery (CERAD-NP) or the Seoul Neuropsychological Screening Battery (SNSB) verbal learning test [16,17]. Memory impairment was defined as a *z* score of –1 or lower on at least one of the following subtests: an immediate word recall test, a delayed word recall test, or a delayed word recognition part of the verbal learning test in CERAD-NP or SNSB, while maintaining essentially normal functional activities; (3) a Mini Mental State Examination (MMSE) score of 27 or lower [18]; (4) not meeting the diagnostic criteria for dementia; (5) a global Clinical Dementia Rating score of 0.5 [19]; (6) taking a stable dose for at least 12 weeks prior to randomization if prescribed acetylcholinesterase inhibitors or memantine; (7) having a study partner who spends more than 8 hours per week with the participant and agrees to assist with the participant's follow-up and clinical evaluations; (8) possession of a personal smartphone and no difficulties using mobile apps; (9) the ability to make phone calls to a study partner independently using the smartphone; (10) no difficulties reading or writing in Korean; and (11) adequate vision and hearing for participating in the clinical trial.

Participants were excluded from the study if any of the following exclusion criteria were met: (1) a history of transient ischemic attack, stroke, or seizure within the past 12 months; (2) a history of severe psychiatric disorders or currently showing unstable psychiatric symptoms; (3) active suicidal intent as assessed by the Columbia-Suicide Severity Rating Scale or a history of treatment for suicidal behavior within the past 5 years [20]; (4) unstable findings upon a physical examination, neurological examination, in vital signs, or the presence of an ongoing unstable physical illness; (5) substance dependence within the past 2 years; (6) the use of prohibited medications as outlined in Table S1 in Multimedia Appendix 1; (7) scheduled to undergo surgery requiring general anesthesia; or (8) participation in any form of cognitive intervention within the past 3 months. Neurologic and psychiatric history, including past substance dependency, current symptoms, physical status, and neurological function, was assessed through clinical interviews conducted by trained psychiatrists and neurologists.

A statistician, independent of the clinical team, carried out the block randomization task, assigning participants to either the study group or the control group at a 1:1 ratio. Throughout the trial, psychiatrists, neurologists, and outcome raters were blinded to the treatment assignment.

The intervention and control groups adhered to an identical visit schedule. Participants completed 5 visits (0-4): screening (visit 0) included screening for study eligibility; week 0 (visit 1, baseline) included baseline demographic data collection and clinical assessments of cognition, function, and quality of life, along with randomization; week 6 (visit 2) investigated compliance with the study protocol and the occurrence of adverse events (AEs); week 12 (visit 3, at the end of the 12-week treatment period) reassessed the same variables from visit 1 and visit 2; and week 24 (visit 4, follow-up assessment 12 weeks after treatment completion) reassessed the same variables from visit 3.

Sample Size Calculation

Sample size was calculated based on a 2-sample proportion test. Delayed recall has been shown to have the highest predictive accuracy (95.2%) in differentiating dementia from MCI [21]. In a previous study, the proportion of participants showing an improvement on delayed recall was 82.8% in the intervention group and 45.16% in the control group [12]. With a 2-sided α of .05 and power of 0.90, and assuming a 1:1 allocation ratio, the minimum sample size per group was determined to be 32 participants. Anticipating a dropout rate of 35%, we intended to enroll 50 participants per group.

Intervention

Study Group: ET-101

Study group participants used ET-101, a mobile app–based MMT program incorporating multimemory strategies. ET-101 is based on an MMT program developed by Youn and colleagues [22] and was produced by Emocog Inc. We previously introduced offline and smart speaker–based versions of this MMT program, and their effectiveness has been well validated [12,14].

Real-time metamemory teaching is a primary feature of ET-101. The training agent of ET-101 interacts with users through voice-based, 2-way communication to enhance metamemory, including meta-knowledge, meta-monitoring, and meta-judgment. Regarding meta-knowledge, on the first day, the agent begins the training by asking users questions about their long-term memory encoding and the memory strategies

they currently use in daily life. Additionally, during the training sessions, the agent introduces multimemory strategies that can enhance long-term memory and explains how these strategies can improve memory performance. This helps users gain objective insight into their cognitive levels and understand how to compensate for cognitive deficits. Regarding meta-monitoring and meta-judgment, the agent explains the daily trends in users' memory spans based on their training results. The agent identifies users' strong and weak memory strategies, helping them recognize their cognitive strengths and weaknesses. Additionally, at the end of the training, the agent provides guidance on how to apply the learned strategies to daily life, facilitating the generalization of memory strategy use. Throughout the metamemory process, the agent provides personalized feedback, adjusts difficulty levels, and modifies training tasks based on users' responses and training results.

ET-101 includes cognitive exercises specifically designed to incorporate memory strategies focusing on attention, imagination, and associations. According to the memory formation process and neuroscience findings, these strategies have been shown to effectively address issues with memory encoding and retrieval commonly observed in older people with cognitive decline [23,24]. Furthermore, attention, imagination, and association processes are known to strengthen related brain regions, including the prefrontal cortex, hippocampus, parahippocampus, precuneus, and the cingulum that connects these areas [23,25]. Attention training includes exercises that enhance trainees' selective focus on stimuli to be memorized, as well as tasks aimed at improving the processing speed and increasing working memory. Imagination training facilitates the transition to long-term memory by linking spatial-temporal memory and visual imagery associated with the objects to be memorized. In association training, 2 or more pieces of information that need to be remembered are connected and assigned mutual meaning, thereby strengthening the semantic cue to aid in long-term memory formation. The detailed content of each activity can be found in Table S2 in Multimedia Appendix 1, and example screenshots are available in Multimedia Appendix 2.

The study group engaged in a 12-week training program using ET-101. The program was structured to involve 2 sessions each day, 7 days a week, lasting approximately 15 minutes per session. The initial session includes 3 representative cognitive exercises, with 1 session each for attention, imagination, and association. In the later session, 4 additional exercises are conducted, selected from 9 available options, excluding the 3 exercises from the initial session, based on a personalized algorithm (Table S3 in Multimedia Appendix 1). A push alarm was provided daily to encourage training.

Control Group: Sham Device

The control group used a sham device over a 12-week period, which featured cognitive assessments and gamification elements but did not include any cognitive training. In the sham device, on the first day, users are asked the same questions as in ET-101 about their long-term memory encoding and the memory strategies they currently use in daily life, presented in a conversational interaction format with the agent. From the second day onward, it offered a gamification element to encourage daily log-ins, in this case displaying a flower that gradually grows. No content related to cognitive training was provided. The guidelines for program usage frequency were identical to those followed by the study group.

Measures

The efficacy of digital therapeutics was evaluated in terms of cognition, ADL, and quality of life. The scales listed below were assessed by a trained psychologist in a face-to-face manner.

Alzheimer's Disease Assessment Scale-Cognitive Subscale-14

ADAS-Cog-14 was used to assess the severity of cognitive dysfunction [26]. ADAS-Cog-14 consists of the following 14 tasks: (1) word recall, (2) commands, (3) constructional praxis, (4) delayed word recall, (5) naming, (6) ideational praxis, (7) orientation, (8) word recognition, (9) maze, (10) number cancellation, (11) remembering instructions, (12) comprehension, (13) word-finding difficulty, and (14) spoken language ability. The total ADAS-Cog-14 score ranges from 0 to 90 points, calculated by summing the number of errors made on each task. A score of 0 indicates no impairment, while a score of 90 reflects the maximum level of impairment. ADAS-Cog subdomains were grouped into 3 domains: memory, language, and praxis [27]. The composite score range for each subdomain is as follows: memory (0-40), language (0-25), and praxis (0-15), with lower scores indicating less impairment.

Korean Mini Mental State Examination, 2nd Edition

MMSE is a brief neuropsychological test that provides an overview of the cognitive function in individuals with cognitive decline [28]. In this study, we used the latest revised version, the Korean Mini Mental State Examination, 2nd Edition (K-MMSE-II), whose validity has already been established [18]. It tests 5 areas of cognitive function: orientation, registration, attention and calculation, recall, and language. The scale ranges from 0 to 30 points, where higher scores indicate less impairment.

Clinical Dementia Rating-Sum of Boxes

Clinical Dementia Rating-Sum of Boxes (CDR-SB) is used to assess the stage of severity in Alzheimer dementia or MCI [29]. The score is determined through interviews with patients and informants, evaluating cognitive functioning across 6 domains: memory, orientation, judgment and problem-solving, community affairs, home and hobbies, and personal care. Each domain is scored from 0 to 3, and the CDR-SB outcome is calculated by summing the scores across all domains, resulting in a total score ranging from 0 to 18. Higher scores indicate greater impairment.

Alzheimer's Disease Composite Score

Alzheimer's Disease Composite Score (ADCOMS) is a tool developed to measure clinical progression and treatment effects in patients with cognitive decline [30]. ADCOMS comprises a total of 12 items, specifically 4 items from ADAS-Cog, 2 items from the MMSE, and all 6 items from CDR-SB. The total score is calculated by applying predetermined weights to each item. Total ADCOMS values range from 0 to 1.97, with higher scores indicating greater impairment.

Digit Symbol Coding

Digit Symbol Coding (DSC) is designed to assess frontal and executive functions [16]. Participants are shown symbols corresponding to numbers from 1 to 9 and are asked to draw the appropriate symbol in 133 blank spaces. Each correctly drawn symbol is awarded 1 point, resulting in a total score ranging from 0 to 133 points. Higher scores indicate better cognitive function.

Clinician Interview–Based Impression of Severity and Clinician Interview–Based Impression of Change Plus Caregiver Input

The Clinician Interview–Based Impression tool is used to assess the severity of cognitive function through semistructured interviews with patients and their caregivers [31]. The Clinician Interview–Based Impression of Severity (CIBIS) is specifically designed to assess baseline severity. Clinicians evaluate responses across 4 domains—general condition, mental or cognitive state, behavior, and ADL—on a scale of 1 to 7 points, where higher scores indicate greater severity. Clinician Interview–Based Impression of Change Plus Caregiver Input (CIBIC-Plus) is used during follow-up to evaluate whether there has been an improvement or decline compared to the baseline level. It assesses 3 aspects: cognition, behavior, and function. Based on these evaluations, the clinician rates the patient's overall condition on a 7-point Likert scale, where 1 indicates very much improved and 7 indicates marked worsening.

Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living

The Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living (ADCS-MCI-ADL) assesses the ability of patients to perform ADL through a structured questionnaire administered to the patient and caregiver by a qualified rater [32]. The scale evaluates both basic and instrumental ADL, with the 24-item version used in this study. Scores range from 0 to 69, with higher scores indicating better ability to independently manage daily living activities, while lower scores reflect a higher degree of dependency when performing such tasks.

EQ-5D-3L and EQ-5D-Visual Analogue Scale

EQ-5D is a tool used to measure health-related quality of life, assessing the 5 dimensions of mobility, self-care, usual activities, pain or discomfort, and anxiety or depression [33]. EQ-5D-3L evaluates each of these dimensions on 3 levels: no problems, some problems, and extreme problems. The responses are then converted into an index score by applying weighted values to each item and summing them, resulting in a score ranging from 0 (death) to 1 (perfect health). In addition, EQ-5D-Visual Analogue Scale (EQ-5D-VAS) captures the respondent's overall assessment of their health on a scale ranging from 0 (worst health imaginable) to 100 (best health imaginable). In this study, both EQ-5D-3L and EQ-5D-VAS were administered to assess the quality of life of patients with MCI as well as the quality of life of their study partners who live with them.

Safety Measures

The occurrence of AEs was monitored actively and systematically, following the Consolidated Standards of Reporting Trials (CONSORT) Harms guideline [34]. AEs were assessed weekly by investigators through telephone contact. Additionally, participants were allowed to report AEs spontaneously at any time during the trial. AEs that involve death, life-threatening conditions, hospitalization or prolonged hospitalization, persistent or significant disability or functional impairment, congenital anomalies or birth defects, or other medically important conditions were classified as serious adverse events (SAEs). Furthermore, the relationship between these AEs and the digital therapeutic device used in this trial was evaluated using a 5-level categorical scale ranging from "Definitely not related" to "Definitely related."

Outcomes

For the primary outcome, participants were categorized as responders or nonresponders based on whether their ADAS-Cog-14 scores were maintained or improved, and the proportion of responders was compared between the ET-101 and sham control groups at weeks 12 and 24. For the secondary outcomes, changes from baseline to 12 and 24 weeks were compared between groups using the following assessments: cognitive function (ADAS-Cog-14, K-MMSE-II, CDR-SB, ADCOMS, and DSC), ADL (CIBIS, CIBIC-Plus, and ADCS-MCI-ADL), and quality of life (EQ-5D-3L and EQ-5D-VAS). For the exploratory analysis, group comparisons of the change scores of the ADAS-Cog-14 subdomains were conducted, including the memory, language, and praxis subdomains. For the safety analysis, the proportions of total AEs, SAEs, and events deemed related to the digital therapeutic intervention were compared between groups.

Statistical Analysis

All statistical analyses were conducted on the full analysis set population, which included all randomized participants who received at least 1 session of the assigned study treatment. This approach is consistent with the modified intention-to-treat concept used in many previous clinical trials [35-37]. Baseline characteristics were compared between the intervention and control groups using an independent 2-tailed *t* test for continuous variables and either a chi-square test or Fisher exact test for categorical variables. The primary outcome, the proportion of responders and nonresponders between the ET-101 group and the control group, was compared by means of a chi-square test. In this study, nonresponder imputation (NRI) was applied, classifying trial dropouts as nonresponders. NRI is a method of handling missing data by assigning all missing cases as nonresponders, thereby preventing treatment benefits from being overestimated. This approach is widely recognized and has been extensively adopted in clinical trials that investigate drug efficacy across various areas [38,39]. In this study, based on NRI, responders were defined as participants who completed the study through week 24 and whose ADAS-Cog-14 score was either maintained or decreased compared to the baseline level, indicating preserved or improved cognitive function [40]. In contrast, nonresponders included those who either failed to

complete the study through week 24 or whose ADAS-Cog-14 score increased compared to the baseline level.

Secondary outcomes, in this case, clinical measures of cognition, ADL, and quality of life, were assessed by analyzing the changes from baseline to week 12 and week 24. Missing data in clinical scales due to trial dropouts were replaced using conditional mean imputation via a regression analysis. For the regression analysis used in imputation, the ADAS-Cog-14, K-MMSE-II, and CDR-SB scores at baseline, as well as the visit and visit-specific values of each clinical measure, were considered. Conditional mean imputation through a regression analysis was conducted using Python (version 3.9.20; Python Software Foundation). Independent *t* tests were used to evaluate whether the score changes differed significantly between the 2 groups. Additionally, a linear mixed model analysis was conducted to examine the treatment-time interaction for each clinical assessment. In this model, treatment, time (indicating baseline, week 12, and week 24), treatment-time interaction, and the baseline score of each scale were included as fixed effects. If a significant difference in the treatment-time interaction between the ET-101 and control groups was found in the linear mixed model, a post hoc analysis was performed to identify the specific time points at which the treatment-time interaction was significant. As a sensitivity analysis, to confirm that the imputation of missing data did not affect the results, the same linear mixed model analysis and post hoc analysis were performed without missing value imputation for variables where a significant treatment-time interaction was found in secondary outcomes.

As an exploratory analysis, the linear mixed model was applied to assess which subdomains of ADAS-Cog-14 (memory, language, and praxis) showed significant treatment-time interaction. If a significant treatment-time interaction was present, a post hoc analysis was conducted to determine the specific time points at which this interaction was significant. As a safety analysis, independent *t* tests were performed for continuous variables, while chi-square tests or Fisher exact tests were used for categorical variables.

All statistical tests were 2-tailed, and *P* scores of <.05 were considered statistically significant. For the post hoc analysis of the linear mixed model, Bonferroni-adjusted *P* scores were reported, with values <.05 indicating statistical significance. All statistical analyses were performed using the *lmerTest* package for the linear mixed model in R software (version 4.4.2; The R Foundation for Statistical Computing).

Ethical Considerations

All participants provided written, informed consent to participate, and their study partners were asked to provide separate written, informed consent. The study was conducted following the Declaration of Helsinki and was approved by the institutional review boards (IRBs) at each of the 7 participating medical centers (Gachon University Gil Medical Center IRB: GCIRB2023-048; Konkuk University Medical Center IRB: 2023-02-017; The Catholic University of Korea, Yeouido St. Mary's Hospital IRB: SC23DDDS0007; Severance Hospital IRB: 10-2023-68; The Catholic University of Korea, Eunpyeong St. Mary's Hospital IRB: PC23DDDS0021; Kangwon National University Hospital IRB: 2023-02-001; SMG-SNU Boramae Medical Center IRB: 1-2023-0005). All participants were anonymized by assigning randomly generated identification numbers. All participants received a compensation of KRW 100,000 (US $68.22) per visit, and study partners were provided with KRW 50,000 (US $34.11) per visit. The study consisted of 5 visits for those who completed the entire study. The reporting of this trial follows the CONSORT statement (Multimedia Appendix 3).

Results

Overview

A total of 140 participants were screened, of whom 40 were excluded based on the inclusion and exclusion criteria (Figure 1). The remaining 100 participants were randomized at a 1:1 ratio. During the clinical trial period, 11 (22.4%) participants in the ET-101 group and 25 (50%) participants in the control group did not complete the study. In the ET-101 group, 1 participant did not use the digital therapeutic device at all after randomization. Therefore, the analysis was conducted on a total of 99 participants in the full analysis set. The dropout rate from the beginning of the training to week 12 was 3 (6%) participants in the ET-101 group and 8 (16%) participants in the sham control group. From week 12 to week 24, the dropout rate was 8 (16%) participants in the ET-101 group and 17 (34%) participants in the sham control group. The adherence rate, defined as the proportion of completed sessions during the 12-week intervention period, was higher in the ET-101 group (mean 83.2%, SD 24.4%) compared to the sham control group (mean 63.8%, SD 31.4%). Daily usage, measured by the number of days participants accessed the program out of the total 84 scheduled training days (12 weeks × 7 days), was also higher in the ET-101 group (mean 69.7, SD 20.2 days) than in the sham control group (mean 53.3, SD 25.8 days). The baseline demographic and clinical characteristics are summarized in Table 1. There were no statistically significant differences in baseline characteristics between the 2 groups.

Figure 1. Flowchart of the study process.

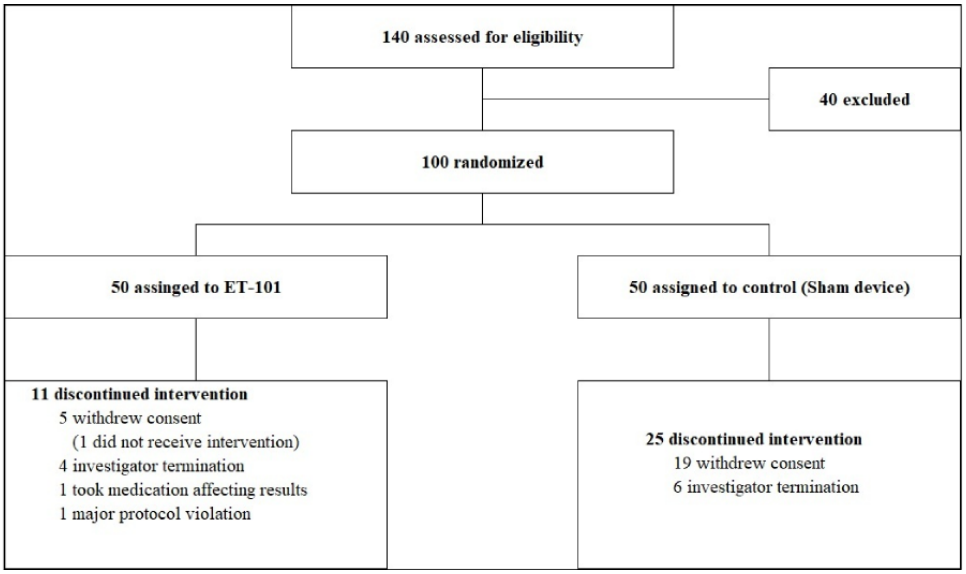


Table 1. Baseline characteristics of study participants.

Characteristics	ET-101 (n=49)	Control (n=50)	*P* value
Age (years), mean (SD)	75 (5.77)	74.8 (5.79)	.84
Sex, n (%)			.91
Male	21 (42.9)	22 (44)	
Female	28 (57.1)	28 (56)	
Education (years), mean (SD)	10.9 (4.47)	10.1 (4.82)	.34
Current medical illness, n (%)			.96
Yes	37 (75.5)	38 (76)	
No	12 (24.5)	12 (24)	
Current use of acetylcholinesterase inhibitors or memantine , n (%)			.90
Yes	18 (36.7)	19 (38)	
No	31 (63.3)	31 (62)	
ADAS-Cog-14[a], mean (SD)[b]	29.2 (6.88)	31 (6.23)	.17
K-MMSE-II[c], mean (SD)	23.9 (2.79)	23 (2.7)	.12
CDR-SB[d], mean (SD)[b]	1.66 (0.856)	2.03 (0.997)	.05
ADCOMS[e], mean (SD)[b]	0.295 (0.146)	0.347 (0.128)	.07
DSC[f], mean (SD)	38.2 (15.4)	34.7 (17.1)	.29
CIBIS[g], mean (SD)[a]	2.67 (0.516)	2.82 (0.523)	.16
ADCS-MCI-ADL[h], mean (SD)	44.3 (6.59)	41.6 (7.06)	.05
EQ-5D-3L patient, mean (SD)	0.929 (0.090)	0.909 (0.087)	.27
EQ-5D-3L study partner, mean (SD)	0.926 (0.092)	0.946 (0.079)	.25
EQ-5D-VAS[i] patient, mean (SD)	78.6 (13)	77.3 (18.7)	.70
EQ-5D-VAS study partner, mean (SD)	76.8 (13.6)	82.2 (14)	.06

[a]ADAS-Cog-14: Alzheimer's Disease Assessment Scale-Cognitive Subscale-14.

[b]Lower scores represent better performance.

[c]K-MMSE-II: Korean Mini Mental State Examination, 2nd Edition.

[d]CDR-SB: Clinical Dementia Rating-Sum of Boxes.

[e]ADCOMS: Alzheimer's Disease Composite Score.

[f]DSC: Digit Symbol Coding.

[g]CIBIS: Clinician Interview–Based Impression of Severity.

[h]ADCS-MCI-ADL: Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living.

[i]EQ-5D-VAS: EQ-5D-Visual Analogue Scale.

Primary Outcome

As shown in Table 2, the proportion of responders and nonresponders did not differ significantly between the ET-101 group and the control group at week 12 (*P*=.47). However, at visit 4, the proportion of responders was significantly higher in the ET-101 group (n=29, 59.2% responders) compared to the sham device group (n=14, 28% responders; *P*=.002).

Table 2. Comparison of the proportion of responders and nonresponders between the ET-101 and control groups.

	ET-101 (n=49), n (%)	Control (n=50), n (%)	*P* value
12 weeks			.47
Responder	20 (40.8)	24 (48)	
Nonresponder	29 (59.2)	26 (52)	
24 weeks			.002
Responder	29 (59.2)	14 (28)	
Nonresponder	20 (40.8)	36 (72)	

Secondary Outcome

The changes in clinical measures from baseline to weeks 12 and 24 are presented in Table 3. For ADAS-Cog-14, the change in scores from baseline to week 12 was compared between the ET-101 and control groups, showing no statistically significant differences (change difference=1.42; *P*=.13). However, at week 24, the change in scores from baseline showed a statistically significant improvement in the ET-101 group compared to the control group (change difference=–2.55; *P*=.02). These findings remained consistent even when examining the treatment-time interaction effect in a linear mixed model that adjusted for baseline ADAS-Cog-14 scores ($F_{2,\ 194}$=7.45; *P*<.001). In the post hoc analysis conducted to identify the specific time points associated with the significant treatment-time interaction, no significant differences were observed between the 2 groups at baseline (estimates=0.022; t_{265}=0.026; Bonferroni-adjusted *P*=.98) or week 12 (estimates=1.44; t_{265}=1.74; Bonferroni-adjusted *P*=.08). However, at week 24, cognitive improvement as measured by ADAS-Cog-14 was statistically significant in the ET-101 group compared to the control group (estimates=–2.53; t_{265}=–3.05; Bonferroni-adjusted *P*=.003; Figure 2). For the other clinical measures, no statistically significant differences were observed.

Table 3. Comparison of cognition, activities of daily living, and quality of life questionnaires between the ET-101 and control groups.

Questionnaire and week	ET-101 (n=49), mean (SD)	Control (n=50), mean (SD)	Change difference[a] (95% CI)	Linear mixed model (treatment-time interaction)	
				F test (*df*)	*P* value
ADAS-Cog-14[b,c]				7.45 (2, 194)	<.001
Week 12	30.3 (7.81)	30.7 (7.40)	1.42 (–0.44 to 3.29)		
Week 24	25.8 (9.52)	30.1 (8.55)	–2.55 (–4.69 to –0.417)[d]		
K-MMSE-II[e]				0.454 (2, 194)	.64
Week 12	24.1 (3.34)	23.6 (2.82)	–0.377 (–1.38 to 0.631)		
Week 24	23.6 (3.16)	23.1 (2.74)	–0.386 (–1.47 to 0.703)		
CDR-SB[b,f]				0.865 (2, 194)	.42
Week 12	1.68 (1.00)	1.92 (1.04)	0.130 (–0.111 to 0.372)		
Week 24	1.43 (1.01)	1.85 (1.02)	–0.055 (–0.381 to 0.272)		
ADCOMS[b,g]				0.608 (2, 194)	.55
Week 12	0.315 (0.175)	0.343 (0.154)	0.024 (–0.014 to 0.062)		
Week 24	0.267 (0.180)	0.310 (0.161)	0.008 (–0.044 to 0.060)		
DSC[h]				0.781 (2, 194)	.46
Week 12	39.1 (14.9)	34.8 (18.8)	0.798 (–1.34 to 2.94)		
Week 24	39.0 (14.1)	33.9 (16.0)	1.62 (–1.29 to 4.52)		
CIBIC-Plus[b,i]				0.526 (1, 97)	.47
Week 12	3.84 (0.717)	4.02 (0.553)	–0.183 (–0.439 to 0.073)[j]		
Week 24	4.02 (0.433)	4.10 (0.364)	–0.080 (–0.239 to 0.080)[j]		
ADCS-MCI-ADL[k]				1.73 (2, 194)	.18
Week 12	43.3 (5.78)	40.7 (6.41)	–0.101 (–1.73 to 1.53)		
Week 24	43.2 (5.86)	41.9 (6.01)	–1.40 to (–3.16 to 0.352)		
EQ-5D-3L patient				0.963 (2, 194)	.38
Week 12	0.925 (0.095)	0.908 (0.098)	–0.004 (–0.038 to 0.030)		
Week 24	0.920 (0.108)	0.925 (0.085)	–0.024 (–0.060 to 0.012)		
EQ-5D-3L study partner				1.50 (2, 194)	.23
Week 12	0.924 (0.101)	0.937 (0.106)	0.008 (–0.028 to 0.045)		
Week 24	0.926 (0.096)	0.967 (0.056)	–0.021 (–0.050 to 0.009)		
EQ-5D-VAS[l] patient				0.705 (2, 194)	.50
Week 12	77.5 (15.9)	75.2 (15.8)	0.978 (–5.26 to 7.21)		
Week 24	72.3 (15.4)	67.5 (13.8)	3.53 (–2.39 to 9.46)		
EQ-5D-VAS study partner				1.92 (2, 194)	.15
Week 12	77.7 (11.6)	81.6 (13.0)	1.38 (–4.14 to 6.90)		
Week 24	77.0 (10.7)	77.2 (12.0)	5.12 (–0.873 to 11.1)		

[a]The score change from baseline to weeks 12 and 24 in the ET-101 group minus the score change from baseline to weeks 12 and 24 in the control group.

[b]Lower scores represent better performance.

[c]ADAS-Cog-14: Alzheimer's Disease Assessment Scale-Cognitive Subscale-14.

[d]*P*<.05.

[e]K-MMSE-II: Korean Mini Mental State Examination, 2nd Edition.

[f]CDR-SB: Clinical Dementia Rating-Sum of Boxes.

[g]ADCOMS: Alzheimer's Disease Composite Score.

[h]DSC: Digit Symbol Coding.

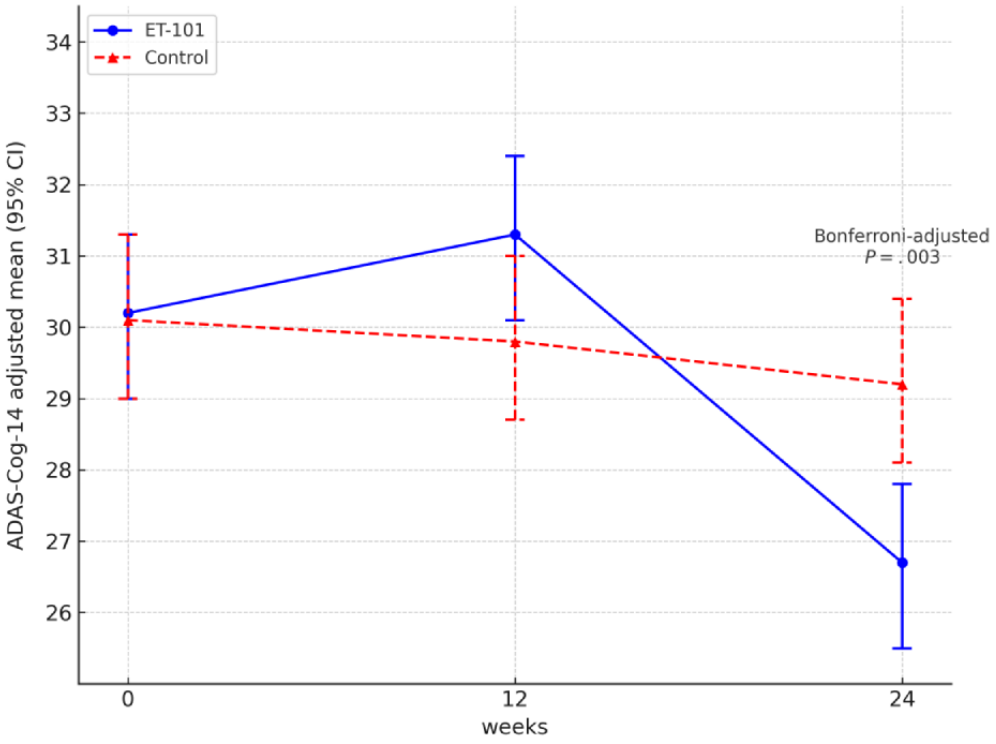
[i]CIBIC-Plus: Clinician Interview–Based Impression of Change Plus Caregiver Input.

[j]Due to the inability to present mean change, the difference in mean change between the groups was replaced with score differences across time points.

[k]ADCS-MCI-ADL: Alzheimer's Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living.

[l]EQ-5D-VAS: EQ-5D-Visual Analogue Scale.

Figure 2. Comparison of adjusted mean change of Alzheimer's Disease Assessment Scale-Cognitive Subscale-14 (ADAS-Cog-14) between ET-101 and the control group using a linear mixed model. Lower scores represent better performance. ADAS-Cog-14 scores adjusted for baseline differences.



Sensitivity Analysis

Even when the linear mixed model was used without applying conditional mean imputation for missing data, the treatment-time interaction remained statistically significant in ADAS-Cog-14 ($F_{2,\ 176}$=3.91; *P*=.02). In the post hoc analysis, no statistically significant differences were observed at baseline (estimates=0.018; t_{232}=0.021; Bonferroni-adjusted *P*=.98) or at week 12 (estimates=1.53; t_{235}=1.69; Bonferroni-adjusted *P*=.09). At week 24, the Bonferroni-adjusted *P* value was .05, indicating a trend toward greater improvement in ADAS-Cog-14 scores in the ET-101 group compared to those of the control group (estimates=–2.13; t_{244}=–1.96).

Exploratory Analysis

The results of the linear mixed model and post hoc analysis for the subdomains of ADAS-Cog-14 are presented in Table S4 in Multimedia Appendix 1. Among the subdomains of ADAS-Cog-14, the treatment-time interaction in the memory domain was statistically significant ($F_{2,\ 194}$=11.2; *P*<.001). In the post hoc analysis, ET-101 showed a statistically significant improvement in the memory domain compared to the control group at week 24 (estimates=–2.50; t_{264}=–4.03; Bonferroni-adjusted *P*<.001).

The treatment-time interaction was also found to be statistically significant for the language domain ($F_{2,\ 194}$=3.99; *P*=.02). In the post hoc analysis, ET-101 demonstrated a statistically significant improvement compared to the control group at week 24 (estimates=–0.807; t_{290}=–3.68; Bonferroni-adjusted *P*<.001). For the praxis subdomain, representing executive function, the treatment-time interaction was not statistically significant ($F_{2,\ 194}$=0.497; *P*=.67).

Safety Analysis

In the ET-101 group, 6 AEs occurred, while 4 AEs were reported in the control group, with no statistically significant difference in AE incidence between the groups (*P*=.53; Table 4). Among the AEs in the ET-101 group, 2 were classified as SAEs due to prolonged hospitalization (non-ST elevation myocardial infarction and herniated disc disease of the lumbar spine); however, the incidence of SAEs also did not differ significantly between the 2 groups (*P*=.47). All reported AEs were determined to be definitely not related to ET-101.

Table 4. Comparison of safety outcomes between ET-101 and the control group.

	ET-101 (n=49), n (%)	Control (n=50), n (%)	*P* value
Serious adverse events			.47
Yes	2 (4.1)	0 (0)	
No	4 (8.2)	4 (8)	
Relevance to digital therapeutics			N/A[a]
Definitely related	0 (0)	0 (0)	
Probably related	0 (0)	0 (0)	
Possibly related	0 (0)	0 (0)	
Probably not related	0 (0)	0 (0)	
Definitely not related	6 (12.2)	4 (8)	
Unknown	0 (0)	0 (0)	
Total	6 (12.2)	4(8)	.52

[a]N/A: not applicable.

Discussion

Principal Findings

This study investigated the efficacy and safety of ET-101 in patients with MCI compared to a sham device control group. We found that the ET-101 group showed significantly greater improvement in the ADAS-Cog-14 score in 24 weeks compared to the control group. Furthermore, no AEs related to the use of ET-101 were observed. Among the various types of digital cognitive training, to the best of our knowledge, this study is the first to investigate the effects of MMT using multimemory strategies in the form of a mobile app.

ET-101, a metamemory-based multistrategic cognitive training program, showed improvement in global cognitive function as well as memory function. In the primary outcome, the responder rate at 24 weeks was significantly higher in the ET-101 group. Additionally, in the secondary outcomes, the ADAS-Cog-14 score showed a statistically significant improvement in the ET-101 group compared to the control group at 24 weeks. According to a meta-analysis that included trials on computerized cognitive training for individuals with MCI, independently delivered digital cognitive training programs, such as ET-101, demonstrated a significant effect on verbal episodic memory, with a standardized mean difference of 0.21 (95% CI 0.04-0.38). This is consistent with the observed memory improvement in this study [41]. ET-101 demonstrated cognitive improvement comparable to that observed with donepezil in terms of ADAS-Cog score changes, suggesting that ET-101 may have a similar cognitive improvement effect [42]. One of the most well-known protective factors against memory decline in MCI is cognitive reserve, which refers to the ability to use cognitive processes and brain networks adaptively to compensate for deterioration [43]. In MCI, there is not only a reduction in the hippocampal volume but also a weakening of the connectivity among key brain regions involved in memory function, such as the frontal, temporal, and parietal lobes [44]. With ET-101, MMT enables individuals to monitor their own memory function objectively and select the most efficient memory strategies tailored to their needs, facilitating the frontal lobe [14]. Additionally, efficient memory strategies, including attention, imagination, and association, enhance the connectivity of brain regions related to cognitive reserve [23,24]. Biological evidence also supports the cognitive reserve–enhancing effects of metamemory. A previous study of MMT found reduced mean diffusivity in the left superior longitudinal fasciculus and corona radiata, which link the frontal, temporal, and parietal lobes [14], suggesting that ET-101 may contribute to enhancing cognitive reserve, which in turn could support improvements in memory function.

Another characteristic of ET-101's cognitive function improvement is that it has a transfer effect on the language domain. In the subdomain analysis of ADAS-Cog-14, the ET-101 group showed a significantly greater improvement in the language domain compared to the control group at 24 weeks. Meta-analyses of several studies of cognitive training in patients with MCI have not shown significant improvements in the language domain [41]. Earlier work has suggested that cognitive training gains typically reflect the training content and that insufficient training on other cognitive domains may therefore make improvements beyond the memory domain less observable [41]. However, MMT-based multistrategic cognitive training, on which ET-101 is based, has consistently demonstrated functional improvements not only in memory but also in the language domain across multiple studies [11,12,14]. One possible explanation for this transfer effect is that the metamemory and memory strategies of ET-101 repeatedly strengthen the connection between episodic memory and semantic memory. MMT helps individuals select compensatory strategies to address their cognitive deficits, and ET-101, in particular, repeatedly trains strategies such as imagination and association, facilitating the connection between episodic and semantic memory [45]. This is especially relevant to lexical production, which is impaired in MCI and requires effortful semantic memory processing, such as in fluency and naming tasks [46]. Notably, naming and semantic fluency have been identified in the literature as strong predictors of the conversion from MCI to Alzheimer dementia, and ET-101 appears to

reinforce cognitive functions that are closely associated with this deterioration [21]. Additional long-term studies are needed to investigate the effects of ET-101 on the rate of conversion from MCI to Alzheimer dementia and to assess its long-term efficacy further.

The observed improvement in the memory and language domains following ET-101 training is closely linked to the memory strategies embedded in the program, which are known to support memory formation. ET-101 primarily incorporates 3 core memory strategies: attention, imagination, and association. First, attention serves as an initial step in memory formation by facilitating selective encoding. Numerous neuroimaging studies have demonstrated that attention enhances memory by activating interactions between the hippocampus and fronto-parietal regions within the dorsal attention network [47-49]. This activation is thought to strengthen neocortical representations, thereby supporting long-term memory storage [50]. Clinically, studies have shown that increased activation of the dorsal attention network is positively correlated with the accuracy and strength of memory recall [51]. Next, imagination shares common neural circuits with episodic memory, involving the hippocampus, medial prefrontal cortex, and anterolateral temporal cortex—regions that are closely associated with memory consolidation [52,53]. Hippocampal replay facilitates the reconstruction of sensory experiences by integrating visual, sensory, and semantic information [52]. This process allows imagination to contribute to memory formation and strengthening, even in the absence of direct sensory input [52]. Regarding association, it has been suggested that newly learned words can facilitate the recall of previously acquired information through retroactive facilitation [54]. Notably, neural evidence indicates that while preexisting semantic relationships are primarily represented within the neocortex, unrelated word pairs can be bound by the hippocampus through associative processes, along with their episodic list context and novelty [54]. Taken together, findings from previous studies suggest that the strategies used in ET-101—attention, imagination, and association—effectively activate the hippocampus and related regions, thereby potentially compensating for the memory deficits observed in individuals with MCI.

However, ET-101 did not show a significant difference compared to the control group on the praxis subdomain, ADL, or quality of life measures. There was no statistically significant difference between the ET-101 and control groups in ADL assessments using CIBIC-Plus and ADCS-MCI-ADL, or in quality-of-life assessments for patients with MCI and their study partners using EQ-5D. Additionally, the praxis domain of ADAS-Cog-14 did not show a significant difference between the groups up to 24 weeks. These findings appear to be related to the characteristics of MCI and the measured scales. MCI involves cognitive decline, but the functional impairment at this stage is not severe enough to warrant a diagnosis of overt dementia [55,56]. Given that those with MCI typically manage daily activities well, ADL and quality-of-life scales may have a ceiling effect, reducing their sensitivity to detect meaningful changes [57]. Therefore, to observe subtle changes in ADL among individuals with MCI, this study used the ADCS-MCI-ADL scale, which is sensitive to detecting changes in ADL among individuals with MCI [32]. In prior pharmacological trials, reported improvements in ADCS-MCI-ADL included participants with both MCI and mild Alzheimer disease [3,58]. However, such improvements in ADCS-MCI-ADL have typically been observed over longer durations, such as 18-month follow-up periods, whereas the 24-week follow-up in our study may not have been long enough to detect meaningful changes [3,58]. This suggests that longer-term follow-up may be necessary to capture significant improvements in ADL or quality of life associated with ET-101. Similarly, in ADAS-Cog-14, the praxis domain has been shown to be less effective in distinguishing different stages of cognitive decline compared to the memory and language domains, and it appears to exhibit a ceiling effect in MCI [26]. Given recent efforts to develop more sensitive scales for detecting changes in ADL, quality of life, and cognitive function in MCI, future studies using these updated tools will be essential for accurately evaluating the effects of ET-101.

The improvement in cognitive function with ET-101 followed a delayed pattern, with significant effects emerging at the 24-week follow-up. While no group differences were observed immediately after the 12-week training period in either the primary or secondary outcomes, the ET-101 group showed significantly greater responder rates and ADAS-Cog-14 improvements at 24 weeks. These findings suggest that the cognitive benefits of ET-101 may not be immediate but instead become evident and persist after the intervention ends. This pattern contrasts with traditional cognitive training, which often shows immediate but short-lived effects. For example, the Advanced Cognitive Training for Independent and Vital Elderly trial reported improvements immediately after memory training, but these effects were not maintained at the 1-year follow-up [7]. In a study of computerized cognitive training that included follow-up data, the initially significant improvements in attention and memory observed immediately after the intervention were no longer present at the 3-month follow-up [8]. Previous meta-analyses have suggested that the effects of memory strategy training tend to be short-lived [59]. Moreover, most studies have not evaluated maintenance effects, resulting in limited evidence on whether training-induced improvements observed during the intervention are sustained in real-life situations [15]. In contrast, ET-101 showed significant gains after the training ended. This delayed onset and sustained improvement align with findings from strategy-based training programs, such as those including memory education, skill practice, and cognitive restructuring of memory-related beliefs [60-62]. Similar to ET-101, these studies did not show immediate cognitive improvements posttraining but demonstrated significant effects during follow-up assessments [60-62]. Such outcomes suggest that benefits emerge gradually as participants apply learned strategies in real-world contexts [61-63]. From a neurobiological perspective, memory consolidation involves synaptic consolidation, which includes short-term molecular processes such as protein synthesis and synaptic potentiation occurring within minutes to hours after memory formation, followed by systems consolidation, during which the interaction between the hippocampus and neocortex is gradually strengthened [64,65]. System consolidation can take several months to years in humans, and sufficient time may

be required to detect changes in memory ensembles and their associated engram networks, which could explain the delayed onset of cognitive improvement observed in this study [65]. However, alternative explanations cannot be excluded. The cognitive improvement observed at the 24-week follow-up may also be partially influenced by practice effects in cognitive testing or by reduced psychological distress after the completion of trial participation, leading to more apparent posttrial improvements. Moreover, the underlying mechanisms of the delayed and lasting cognitive effects remain unclear, and further longitudinal follow-up and functional neuroimaging studies are warranted to elucidate the temporal dynamics of training-induced neural plasticity. Future meta-analyses and systematic reviews should also investigate the characteristics of cognitive training protocols associated with either immediate or delayed effects to clarify the factors that determine these distinct temporal response patterns.

The cognitive improvement effect of ET-101 was observed in the ADAS-Cog-14 case, but not on other cognitive function assessment scales. In the secondary outcomes, the measures K-MMSE-II, CDR-SB, ADCOMS, and DSC did not show a statistically significant treatment-time interaction. These findings may be related to the sensitivity of cognitive function assessment scales. Previous studies comparing ADAS-Cog-14 with other cognitive measures have shown that ADAS-Cog can detect cognitive changes in MCI significantly more sensitively compared to MMSE and CDR-SB [66,67]. Notably, this study used the 14-item version of ADAS-Cog, which has been specifically reported to be highly sensitive for MCI assessment [26]. Therefore, the observed improvement in ADAS-Cog-14, but not on other scales, is likely due to differences in the sensitivity of these assessments in MCI, with ADAS-Cog-14 being more capable of detecting subtle cognitive changes.

Limitations

One limitation of this study is the relatively high number of dropouts during the trial. During the 12-week training period, 3 (6%) participants in the ET-101 group and 8 (16%) participants in the sham control group dropped out, values that fall within or below the typical range of dropout rates (30%-40%) reported in previous digital intervention studies [68,69]. However, between week 12 and week 24—after the intervention had ended—additional dropouts occurred in both groups: 8 (16%) participants in the ET-101 group and 17 (34%) in the sham control group. This increase may reflect reduced motivation to continue participation in the absence of active training. As these attrition rates necessitated an imputation strategy for missing data, the interpretation of the results should be approached with caution. In particular, the NRI method may overpenalize attrition and potentially bias the treatment effect estimates downward, especially when the reasons for dropout differ between groups. Therefore, to minimize the impact of the imputation used in this study, we conducted analyses to verify whether the results remained consistent across different statistical approaches. These included a comparison of the proportion of participants showing maintenance or improvement in cognitive function between the ET-101 and sham control groups based on NRI, as well as analyses of continuous cognitive scores using both independent *t* tests and linear mixed models. Furthermore, to minimize the impact of the imputation as conducted here, we used the same linear mixed model without imputation for missing data. Across these multiple analytic approaches, we consistently observed that ET-101 produced significantly better maintenance or improvement of cognitive function compared to sham controls at the 24-week follow-up.

Additionally, certain clinical variables, in this case CDR-SB and EQ-5D-VAS, had *P* values close to .05 in baseline comparisons, indicating that the 2 groups were not entirely homogeneous in terms of certain variables. To address this, we incorporated each scale's baseline score as a fixed effect in the linear mixed model when analyzing secondary outcomes. This approach allowed us to confirm that the treatment-time interaction remained statistically significant even after adjusting for these differences.

Third, although participants were not explicitly informed of their group assignment, differences in content between the ET-101 and sham control groups may have caused some participants to infer their allocation based on the app experience. This is a common issue in studies evaluating the efficacy of software-based interventions, where a sham condition must be provided without including the active training components. As such, there is a potential risk of performance and expectation bias between groups, which must be taken into consideration when interpreting the results.

Fourth, this study included only participants who were able to own and operate a smartphone, which may have led to the inclusion of a relatively more tech-savvy population. However, this inclusion criterion was necessary because ET-101 is designed to be used independently without additional assistance during the training process, targeting individuals with basic smartphone literacy. In particular, in the context of Korea, epidemiological data from 2023 show that approximately 95% of individuals aged 60 years and older own a personal mobile phone [70]. Therefore, this inclusion criterion does not substantially deviate from the characteristics of the older adult population in Korea.

Conclusions

Metamemory-based multistrategic cognitive training, ET-101, was found to improve cognitive function compared to a sham device control group. This effect was observed in both the targeted domain of memory and in the language domain, indicating a transfer effect. Additionally, the cognitive improvement became pronounced at 24 weeks, suggesting that ET-101 has a sustained effect beyond the training period. These findings indicate that ET-101 has the potential to provide effective MMT to a broader population with MCI, overcoming location and personnel limitations via its mobile app–based platform.

Acknowledgments

This research was financially supported by the Ministry of Trade, Industry, and Energy, Korea, under the “Regional Innovation Cluster Development Program (R&D, P0025292)” supervised by the Korea Institute for Advancement of Technology. Additionally, ChatGPT 4o was used solely for the purpose of grammar checking. No AI tools were used to generate original content or interpret study findings.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

KHP conceived the study, designed the research, and supervised the project. KHP, KYK, EK, HJJ, DHP, HKL, YS, JWJ, and YK collected and organized patient data. SK conducted the data analysis and drafted the manuscript. SK, JIL, and LS interpreted the results. SL was involved in drafting the description of the ET-101. All authors critically reviewed and revised the manuscript. All authors have read and approved the final version for submission.

Conflicts of Interest

Emocog Inc owns the mobile app evaluated in this study and has a commercial interest in its future distribution. SL is employed as a software developer at Emocog Inc. SK was employed in the Research and Development team at Emocog Inc. until August 31, 2025. LS is the Chief Operating Officer of Cogthera GmbH, a subsidiary of Emocog Inc. Cogthera GmbH is responsible for the commercialization and planned distribution of a modified version of this application in Germany. However, SK and LS were not involved in participant enrollment or data collection, which was conducted independently by other authors (KYK, EK, HJJ, DHP, HKL, YS, JWJ, YK, and KHP), who had no conflicts of interest.

Multimedia Appendix 1
Supplementary tables.
[XLSX File (Microsoft Excel File), 16 KB - mhealth_v14i1e73464_app1.xlsx]

Multimedia Appendix 2
Sample exercise from the ET-101 cognitive training program.
[PDF File (Adobe PDF File), 413 KB - mhealth_v14i1e73464_app2.pdf]

Multimedia Appendix 3
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1148 KB - mhealth_v14i1e73464_app3.pdf]

References

1. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. Lancet Public Health 2022;7(2):e105-e125 [FREE Full text] [doi: 10.1016/S2468-2667(21)00249-8] [Medline: 34998485]
2. Xu Z, Sun W, Zhang D, Chung VC, Sit RW, Wong SY. Comparative effectiveness of interventions for global cognition in patients with mild cognitive impairment: a systematic review and network meta-analysis of randomized controlled trials. Front Aging Neurosci 2021;13:653340 [FREE Full text] [doi: 10.3389/fnagi.2021.653340] [Medline: 34220484]
3. van Dyck CH, Swanson CJ, Aisen P, Bateman RJ, Chen C, Gee M, et al. Lecanemab in early alzheimer's disease. N Engl J Med 2023;388(1):9-21. [doi: 10.1056/NEJMoa2212948] [Medline: 36449413]
4. Raimo S, Cropano M, Gaita M, Maggi G, Cavallo ND, Roldan-Tapia MD, et al. The efficacy of cognitive training on neuropsychological outcomes in mild cognitive impairment: a meta-analysis. Brain Sci 2023;13(11):1510 [FREE Full text] [doi: 10.3390/brainsci13111510] [Medline: 38002471]
5. Chapman SB, Aslan S, Spence JS, Hart JJ, Bartz EK, Didehbani N, et al. Neural mechanisms of brain plasticity with complex cognitive training in healthy seniors. Cereb Cortex 2015;25(2):396-405 [FREE Full text] [doi: 10.1093/cercor/bht234] [Medline: 23985135]
6. Hertzog C, Kramer AF, Wilson RS, Lindenberger U. Enrichment effects on adult cognitive development: can the functional capacity of older adults be preserved and enhanced? Psychol Sci Public Interest 2008;9(1):1-65. [doi: 10.1111/j.1539-6053.2009.01034.x] [Medline: 26162004]
7. Sisco SM, Marsiske M, Gross AL, Rebok GW. The influence of cognitive training on older adults' recall for short stories. J Aging Health 2013;25(8 Suppl):230S-248S [FREE Full text] [doi: 10.1177/0898264313501386] [Medline: 24385636]

8. Zelinski EM, Spina LM, Yaffe K, Ruff R, Kennison RF, Mahncke HW, et al. Improvement in memory with plasticity-based adaptive cognitive training: results of the 3-month follow-up. J Am Geriatr Soc 2011;59(2):258-265. [doi: 10.1111/j.1532-5415.2010.03277.x] [Medline: 21314646]
9. Flavell JH. Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. Am Psychol 1979;34(10):906-911. [doi: 10.1037/0003-066x.34.10.906]
10. Tarricone P. The Taxonomy of Metacognition. London: Psychology Ppress; 2011.
11. Youn JH, Lee JY, Kim S, Ryu SH. Multistrategic memory training with the metamemory concept in healthy older adults. Psychiatry Investig 2011;8(4):354-361 [FREE Full text] [doi: 10.4306/pi.2011.8.4.354] [Medline: 22216046]
12. Kim J, Shin E, Han K, Park S, Youn JH, Jin G, et al. Efficacy of smart speaker-based metamemory training in older adults: case-control cohort study. J Med Internet Res 2021;23(2):e20177 [FREE Full text] [doi: 10.2196/20177] [Medline: 33591276]
13. Sella E, Carbone E, Vincenzi M, Toffalini E, Borella E. Efficacy of memory training interventions targeting metacognition for older adults: a systematic review and meta-analysis. Aging Ment Health 2023;27(4):674-694. [doi: 10.1080/13607863.2022.2122931] [Medline: 36218025]
14. Youn JH, Ryu SH, Lee JY, Park S, Cho SJ, Kwon H, et al. Brain structural changes after multi-strategic metamemory training in older adults with subjective memory complaints: a randomized controlled trial. Brain Behav 2019;9(5):e01278 [FREE Full text] [doi: 10.1002/brb3.1278] [Medline: 30916450]
15. Chan ATC, Ip RTF, Tran JYS, Chan JYC, Tsoi KKF. Computerized cognitive training for memory functions in mild cognitive impairment or dementia: a systematic review and meta-analysis. NPJ Digit Med 2024;7(1):1 [FREE Full text] [doi: 10.1038/s41746-023-00987-5] [Medline: 38172429]
16. Ryu HJ, Yang DW. The Seoul Neuropsychological Screening Battery (SNSB) for comprehensive neuropsychological assessment. Dement Neurocogn Disord 2023;22(1):1-15 [FREE Full text] [doi: 10.12779/dnd.2023.22.1.1] [Medline: 36814700]
17. Lee JH, Lee KU, Lee DY, Kim KW, Jhoo JH, Kim JH, et al. Development of the Korean version of the Consortium to Establish a Registry for Alzheimer's Disease Assessment Packet (CERAD-K): clinical and neuropsychological assessment batteries. J Gerontol B Psychol Sci Soc Sci 2002;57(1):P47-P53. [doi: 10.1093/geronb/57.1.p47] [Medline: 11773223]
18. Baek MJ, Kim K, Park YH, Kim S. The validity and reliability of the Mini-Mental State Examination-2 for detecting mild cognitive impairment and Alzheimer's disease in a Korean population. PLoS One 2016;11(9):e0163792 [FREE Full text] [doi: 10.1371/journal.pone.0163792] [Medline: 27668883]
19. Choi SH, Na DL, Lee BH, Hahm DS, Jeong JH, Yoon SJ, et al. Estimating the validity of the Korean version of expanded Clinical Dementia Rating (CDR) scale. J Korean Neurol Assoc 2001;19(6):585-591 [FREE Full text]
20. Pai D, Woo JM, Son MH, Lee C. The reliability and validity of the Korean version of Columbia-Suicide Severity Rating Scale in alcohol dependent patients. J Korean Neuropsychiatr Assoc 2015;54(2):222-227 [FREE Full text]
21. Belleville S, Fouquet C, Hudon C, Zomahoun HTV, Croteau J, Consortium for the Early Identification of Alzheimer's disease-Quebec. Neuropsychological measures that predict progression from mild cognitive impairment to Alzheimer's type dementia in older adults: a systematic review and meta-analysis. Neuropsychol Rev 2017;27(4):328-353 [FREE Full text] [doi: 10.1007/s11065-017-9361-5] [Medline: 29019061]
22. Youn J, Park S, Lee J, Cho S, Kim J, Ryu S. Cognitive improvement in older adults with mild cognitive impairment: evidence from a multi-strategic metamemory training. J Clin Med 2020;9(2):362 [FREE Full text] [doi: 10.3390/jcm9020362] [Medline: 32013035]
23. Cabeza R, Grady CL, Nyberg L, McIntosh AR, Tulving E, Kapur S, et al. Age-related differences in neural activity during memory encoding and retrieval: a positron emission tomography study. J Neurosci 1997;17(1):391-400 [FREE Full text] [doi: 10.1523/JNEUROSCI.17-01-00391.1997] [Medline: 8987764]
24. Craik FIM, Rose NS. Memory encoding and aging: a neurocognitive perspective. Neurosci Biobehav Rev 2012;36(7):1729-1739. [doi: 10.1016/j.neubiorev.2011.11.007] [Medline: 22155274]
25. Hampstead BM, Stringer AY, Stilla RF, Sathian K. Mnemonic strategy training increases neocortical activation in healthy older adults and patients with mild cognitive impairment. Int J Psychophysiol 2020;154:27-36 [FREE Full text] [doi: 10.1016/j.ijpsycho.2019.04.011] [Medline: 31067489]
26. Kueper JK, Speechley M, Montero-Odasso M. The Alzheimer's Disease Assessment Scale-Cognitive Subscale (ADAS-Cog): modifications and responsiveness in pre-dementia populations. A narrative review. J Alzheimers Dis 2018;63(2):423-444 [FREE Full text] [doi: 10.3233/JAD-170991] [Medline: 29660938]
27. Verma N, Beretvas SN, Pascual B, Masdeu JC, Markey MK, Alzheimer’s Disease Neuroimaging Initiative. New scoring methodology improves the sensitivity of the Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) in clinical trials. Alzheimers Res Ther 2015;7(1):64 [FREE Full text] [doi: 10.1186/s13195-015-0151-0] [Medline: 26560146]
28. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12(3):189-198. [doi: 10.1016/0022-3956(75)90026-6] [Medline: 1202204]
29. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982;140:566-572. [doi: 10.1192/bjp.140.6.566] [Medline: 7104545]

30. Tahami Monfared AA, Houghton K, Zhang Q, Mauskopf J, Alzheimer's Disease Neuroimaging Initiative. Staging disease severity using the Alzheimer's disease composite score (ADCOMS): a retrospective data analysis. Neurol Ther 2022;11(1):413-434 [FREE Full text] [doi: 10.1007/s40120-022-00326-y] [Medline: 35099758]
31. Knopman DS, Knapp MJ, Gracon SI, Davis CS. The Clinician Interview-Based Impression (CIBI): a clinician's global change rating scale in Alzheimer's disease. Neurology 1994;44(12):2315-2321. [doi: 10.1212/wnl.44.12.2315] [Medline: 7991118]
32. Pedrosa H, De Sa A, Guerreiro M, Maroco J, Simoes MR, Galasko D, et al. Functional evaluation distinguishes MCI patients from healthy elderly people--the ADCS/MCI/ADL scale. J Nutr Health Aging 2010;14(8):703-709 [FREE Full text] [doi: 10.1007/s12603-010-0102-1] [Medline: 20922349]
33. Devlin N, Parkin D, Janssen B. Methods for Analysing and Reporting EQ-5D Data. Cham (CH): Springer; 2020.
34. Junqueira DR, Zorzela L, Golder S, Loke Y, Gagnier JJ, Julious SA, CONSORT Harms Group. CONSORT Harms 2022 statement, explanation, and elaboration: updated guideline for the reporting of harms in randomised trials. BMJ 2023;381:e073725 [FREE Full text] [doi: 10.1136/bmj-2022-073725] [Medline: 37094878]
35. Kahan BC, White IR, Edwards M, Harhay MO. Using modified intention-to-treat as a principal stratum estimator for failure to initiate treatment. Clin Trials 2023;20(3):269-275 [FREE Full text] [doi: 10.1177/17407745231160074] [Medline: 36916466]
36. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ 2010;340:c2697 [FREE Full text] [doi: 10.1136/bmj.c2697] [Medline: 20547685]
37. Gupta SK. Intention-to-treat concept: a review. Perspect Clin Res 2011;2(3):109-112 [FREE Full text] [doi: 10.4103/2229-3485.83221] [Medline: 21897887]
38. Floden L, Bell ML. Imputation strategies when a continuous outcome is to be dichotomized for responder analysis: a simulation study. BMC Med Res Methodol 2019;19(1):161 [FREE Full text] [doi: 10.1186/s12874-019-0793-x] [Medline: 31345166]
39. McInnes IB, Asahina A, Coates LC, Landewé R, Merola JF, Ritchlin CT, et al. Bimekizumab in patients with psoriatic arthritis, naive to biologic treatment: a randomised, double-blind, placebo-controlled, phase 3 trial (BE OPTIMAL). Lancet 2023;401(10370):25-37 [FREE Full text] [doi: 10.1016/S0140-6736(22)02302-9] [Medline: 36493791]
40. Rockwood K, Howlett SE, Hoffman D, Schindler R, Mitnitski A. Clinical meaningfulness of Alzheimer's Disease Assessment Scale-Cognitive Subscale change in relation to goal attainment in patients on cholinesterase inhibitors. Alzheimers Dement 2017;13(10):1098-1106. [doi: 10.1016/j.jalz.2017.02.005] [Medline: 28341540]
41. Hill NTM, Mowszowski L, Naismith SL, Chadwick VL, Valenzuela M, Lampit A. Computerized cognitive training in older adults with mild cognitive impairment or dementia: a systematic review and meta-analysis. Am J Psychiatry 2017;174(4):329-340. [doi: 10.1176/appi.ajp.2016.16030360] [Medline: 27838936]
42. Doody RS, Ferris SH, Salloway S, Sun Y, Goldman R, Watkins WE, et al. Donepezil treatment of patients with MCI: a 48-week randomized, placebo-controlled trial. Neurology 2009;72(18):1555-1561. [doi: 10.1212/01.wnl.0000344650.95823.03] [Medline: 19176895]
43. Corbo I, Marselli G, Di Ciero V, Casagrande M. The protective role of cognitive reserve in mild cognitive impairment: a systematic review. J Clin Med 2023;12(5):1759 [FREE Full text] [doi: 10.3390/jcm12051759] [Medline: 36902545]
44. Ries ML, Carlsson CM, Rowley HA, Sager MA, Gleason CE, Asthana S, et al. Magnetic resonance imaging characterization of brain structure and function in mild cognitive impairment: a review. J Am Geriatr Soc 2008;56(5):920-934 [FREE Full text] [doi: 10.1111/j.1532-5415.2008.01684.x] [Medline: 18410325]
45. Abraham A, Bubic A. Semantic memory as the root of imagination. Front Psychol 2015;6:325 [FREE Full text] [doi: 10.3389/fpsyg.2015.00325] [Medline: 25852626]
46. Jokel R, Lima BS, Fernandez A, Murphy KJ. Language in amnestic mild cognitive impairment and dementia of Alzheimer's type: quantitatively or qualitatively different? Dementia and geriatric cognitive disorders extra. Dement Geriatr Cogn Disord Extra 2019;9(1):136-151 [FREE Full text]
47. Geng F, Xu W, Riggins T. Interactions between the hippocampus and fronto-parietal regions during memory encoding in early childhood. Hippocampus 2022;32(2):108-120. [doi: 10.1002/hipo.23380] [Medline: 34329507]
48. Buckner RL, Kelley WM, Petersen SE. Frontal cortex contributes to human memory formation. Nat Neurosci 1999;2(4):311-314. [doi: 10.1038/7221] [Medline: 10204536]
49. Chun MM, Turk-Browne NB. Interactions between attention and memory. Curr Opin Neurobiol 2007;17(2):177-184. [doi: 10.1016/j.conb.2007.03.005] [Medline: 17379501]
50. Kim H. Attention-versus significance-driven memory formation: taxonomy, neural substrates, and meta-analyses. Neurosci Biobehav Rev 2022;138:104685. [doi: 10.1016/j.neubiorev.2022.104685] [Medline: 35526692]
51. Das A, Menon V. Hippocampal-parietal cortex causal directed connectivity during human episodic memory formation: replication across three experiments. bioRxiv Preprint posted online on March 21, 2024 [FREE Full text] [doi: 10.1101/2023.11.07.566056] [Medline: 37986855]
52. Spens E, Burgess N. A generative model of memory construction and consolidation. Nat Hum Behav 2024;8(3):526-543 [FREE Full text] [doi: 10.1038/s41562-023-01799-z] [Medline: 38242925]

53. Huang Q, Xiao Z, Yu Q, Luo Y, Xu J, Qu Y, et al. Replay-triggered brain-wide activation in humans. Nat Commun 2024;15(1):7185 [FREE Full text] [doi: 10.1038/s41467-024-51582-5] [Medline: 39169063]
54. Antony JW, Romero A, Vierra AH, Luenser RS, Hawkins RD, Bennion KA. Semantic relatedness retroactively boosts memory and promotes memory interdependence across episodes. Elife 2022;11:e72519 [FREE Full text] [doi: 10.7554/eLife.72519] [Medline: 35704025]
55. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5). Washington, DC: American Psychiatric Association; 2013.
56. Schneider LS, Goldberg TE. Composite cognitive and functional measures for early stage Alzheimer's disease trials. Alzheimers Dement (Amst) 2020;12(1):e12017 [FREE Full text] [doi: 10.1002/dad2.12017] [Medline: 32432155]
57. Gopalakrishnan P, Tiwari S, Nagaraja R, Krishnan G. Quality of life in persons with mild cognitive impairment: a systematic review and meta-analysis. Dement Neuropsychol 2024;18:e20230093 [FREE Full text] [doi: 10.1590/1980-5764-DN-2023-0093] [Medline: 39193465]
58. Budd Haeberlein S, Aisen PS, Barkhof F, Chalkias S, Chen T, Cohen S, et al. Two randomized phase 3 studies of aducanumab in early alzheimer's disease. J Prev Alzheimers Dis 2022;9(2):197-210. [doi: 10.14283/jpad.2022.30] [Medline: 35542991]
59. Hudes R, Rich JB, Troyer AK, Yusupov I, Vandermorris S. The impact of memory-strategy training interventions on participant-reported outcomes in healthy older adults: A systematic review and meta-analysis. Psychol Aging 2019;34(4):587-597. [doi: 10.1037/pag0000340] [Medline: 30896195]
60. Rapp S, Brenes G, Marsh AP. Memory enhancement training for older adults with mild cognitive impairment: a preliminary study. Aging Ment Health 2002;6(1):5-11. [doi: 10.1080/13607860120101077] [Medline: 11827617]
61. Belleville S, Cuesta M, Bier N, Brodeur C, Gauthier S, Gilbert B, et al. Five-year effects of cognitive training in individuals with mild cognitive impairment. Alzheimers Dement (Amst) 2024;16(3):e12626. [doi: 10.1002/dad2.12626] [Medline: 39246830]
62. Yang L. Maintained and delayed benefits of executive function training and low-intensity aerobic exercise over a 3.5-year period in older adults. Front Aging Neurosci 2022;14:905886 [FREE Full text] [doi: 10.3389/fnagi.2022.905886] [Medline: 35847677]
63. Kinsella GJ, Ames D, Storey E, Ong B, Pike KE, Saling MM, et al. Strategies for improving memory: a randomized trial of memory groups for older people, including those with mild cognitive impairment. J Alzheimers Dis 2016;49(1):31-43. [doi: 10.3233/JAD-150378] [Medline: 26444773]
64. Winocur G, Moscovitch M, Sekeres M. Memory consolidation or transformation: context manipulation and hippocampal representations of memory. Nat Neurosci 2007;10(5):555-557. [doi: 10.1038/nn1880] [Medline: 17396121]
65. Quillfeldt JA. Temporal flexibility of systems consolidation and the Synaptic Occupancy/Reset Theory (SORT): cues about the nature of the engram. Front Synaptic Neurosci 2019;11:1 [FREE Full text] [doi: 10.3389/fnsyn.2019.00001] [Medline: 30814946]
66. Balsis S, Benge JF, Lowe DA, Geraci L, Doody RS. How do scores on the ADAS-Cog, MMSE, and CDR-SOB correspond? Clin Neuropsychol 2015;29(7):1002-1009. [doi: 10.1080/13854046.2015.1119312] [Medline: 26617181]
67. Wessels AM, Dowsett SA, Sims JR. Detecting treatment group differences in Alzheimer's disease clinical trials: a comparison of Alzheimer's Disease Assessment Scale-Cognitive subscale (ADAS-Cog) and the Clinical Dementia Rating-Sum of Boxes (CDR-SB). J Prev Alzheimers Dis 2018;5(1):15-20 [FREE Full text] [doi: 10.14283/jpad.2018.2] [Medline: 29405227]
68. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
69. Iqhrammullah M, Yudhistira Refin R, Fitria Andika F, Amirah S, Fahd Abdurrahman M, Alina M, et al. Dropout rate in clinical trials of smartphone apps for diabetes management: a meta-analysis. Diabetes Res Clin Pract 2024;212:111723. [doi: 10.1016/j.diabres.2024.111723] [Medline: 38830484]
70. 2023 survey on broadcasting media usage behavior. Korea Communications Commission. 2023. URL: https://www.kcc.go.kr/user.do?mode=view&page=A05030000&boardId=1113&boardSeq=58975 [accessed 2025-05-12]

Abbreviations

ADAS-Cog: Alzheimer’s Disease Assessment Scale-Cognitive Subscale
ADCOMS: Alzheimer’s Disease Composite Score
ADCS-MCI-ADL: Alzheimer’s Disease Cooperative Study-Mild Cognitive Impairment-Activities of Daily Living
ADL: activities of daily living
AE: adverse event
CDR-SB: Clinical Dementia Rating-Sum of Boxes
CERAD-NP: Consortium to Establish a Registry for Alzheimer’s Disease Neuropsychological Assessment Battery
CIBIC-Plus: Clinician Interview–Based Impression of Change Plus Caregiver Input
CIBIS: Clinician Interview–Based Impression of Severity
CONSORT: Consolidated Standards of Reporting Trials

DSC: Digit Symbol Coding
EQ-5D-VAS: EQ-5D-Visual Analogue Scale
IRB: institutional review board
K-MMSE-II: Korean Mini Mental State Examination, 2nd Edition
MCI: mild cognitive impairment
MMSE: Mini Mental State Examination
MMT: metamemory training
NRI: nonresponder imputation
SAE: serious adverse event
SNSB: Seoul Neuropsychological Screening Battery

Edited by L Buis; submitted 24.Mar.2025; peer-reviewed by KW Turk, K Yuenyongchaiwat; comments to author 01.May.2025; revised version received 19.Nov.2025; accepted 20.Nov.2025; published 19.Jan.2026.

Please cite as:
Kang S, Lim JI, Stenzel L, Kim KY, Kim E, Jeon HJ, Park DH, Lim HK, Shim Y, Jang JW, Kim Y, Lee S, Park KH
Efficacy and Safety of Mobile App–Based Metamemory Cognitive Training for Mild Cognitive Impairment: Multicenter Randomized Clinical Trial
JMIR Mhealth Uhealth 2026;14:e73464
URL: https://mhealth.jmir.org/2026/1/e73464
doi:10.2196/73464
PMID:

Assessing Wearable mHealth Adherence in Underserved Adolescents and its Associations With Physical Activity, Sports, and Safety Perceptions: Prospective Cohort Study

Annabel Nunez-Gaunaurd[1*], PT, PhD; Michele Raya[2*], PT, PhD

[1]Department of Physical Therapy, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, 11200 SW 8 St, AHC3-424, Miami, FL, United States

[2]Department of Physical Therapy, Miller School of Medicine, University of Miami, Coral Gables, FL, United States

[*]all authors contributed equally

Corresponding Author:
Annabel Nunez-Gaunaurd, PT, PhD
Department of Physical Therapy, Nicole Wertheim College of Nursing and Health Sciences, Florida International University, 11200 SW 8 St, AHC3-424, Miami, FL, United States

Abstract

Background: Adolescents from underserved communities, particularly Black and Hispanic youth, engage in lower levels of physical activity (PA), increasing their risk for chronic disease. Conventional interventions often face barriers such as limited access to safe environments. Wearable mobile health technologies offer scalable and context-sensitive solutions; however, predictors of sustained adherence in school-based settings among high-risk populations remain underexplored.

Objective: This study aims to examine the behavioral and contextual predictors of adherence to a consumer-grade wearable PA tracker among underserved high school students.

Methods: In this school-based observational study, 63 students (mean age 14.8, SD 1.17 years) enrolled in physical education received Fitbit devices. Adherence was defined as ≥21 valid days of step count data. Measures included self-reported PA behaviors, neighborhood perceptions, physical fitness (including anthropometrics), and device adherence. Group comparisons were conducted using *t* tests and chi-square tests. Logistic regression was used to identify predictors of adherence.

Results: Overall, 73% (46/63) of participants met the adherence threshold. Adherent students reported fewer days of moderate-to-vigorous PA (2 vs 4 days/week; *P*=.004), lower team sports participation (21/46, 46% vs 12/17, 71%; *P*=.004), and higher perceived neighborhood safety (*P*=.02). In adjusted models, lower PA frequency, greater perceived safety, and neighborhood walkability significantly predicted adherence (χ^2_6=16.23*; P*=.01, Nagelkerke R^2=0.61).

Conclusions: Wearable mobile health technologies show promise for engaging underserved adolescents in PA, particularly those with lower baseline activity and limited access to structured sports. Key predictors of adherence included perceived neighborhood walkability, team sports participation, and prior PA behavior. School-based deployment of wearable devices should emphasize personalized goals and autonomy-supportive strategies to foster sustained engagement and promote PA among high-risk youth.

(JMIR Mhealth Uhealth 2026;14:e80465) doi:10.2196/80465

KEYWORDS

mHealth; physical activity; health promotion; obesity; wearable adherence; functional measures; mobile health; wearables; adolescents; school-based

Introduction

With national and international guidelines recommending at least 60 minutes of moderate-to-vigorous physical activity (MVPA) daily for children and adolescents, physical inactivity remains a pressing public health concern [1,2]. Regular physical activity (PA) is associated with numerous health benefits, including lower resting heart rate (HR), a key indicator of cardiovascular health [3,4], improved functional fitness [5], and enhanced long-term physical and mental well-being [6]. Despite these benefits, only 22% of adolescents meet the recommended PA guidelines [7], with participation rates significantly lower among Black, Hispanic, and immigrant youth [8]. Many children and adolescents in racially diverse and low-resource neighborhoods lack access to consistent and supportive PA opportunities outside school [9]. Furthermore, the growing financial and time demands placed on parents for youth sports participation [10] have made consistent PA opportunities

increasingly inaccessible for many adolescents, particularly those from underserved communities.

Consumer-grade wearable mobile health (mHealth) technologies may offer a scalable, user-friendly means to assess and promote PA engagement [11]. Awareness and use of step-based metrics from wearables can provide an opportunity for self-monitoring, goal setting, and motivation to increase PA behavior and improve health and fitness. For example, typically step-based metrics ranging from 7000 to 11,700 steps per day are associated with meeting Centers for Disease Control and Prevention (CDC)-recommended PA levels [12-14] and with reducing cardiovascular and mortality risks [15]. Conversely, taking fewer than 7000 steps per day may indicate insufficient PA behavior [14]. Wearables can help adolescents become more aware of their PA levels and facilitate tracking of these metrics [16]. There is growing evidence supporting the role of mHealth technologies, including smartphones, apps, and wearables, in promoting positive health behaviors and some improvement in weight status among youth [17]. These devices incorporate behavioral change techniques, such as goal setting, self-monitoring, and feedback [18], and may be particularly useful in school-based programs where environmental constraints can be addressed.

Traditional PA opportunities often rely on structured sports or clinic programs outside of school, which may be inaccessible due to socioeconomic and parental involvement time constraints, limited transportation, and neighborhood safety concerns for underserved youth [19]. Additionally, limited access to safe recreational environments and health-promoting infrastructure further compounds these risks, reinforcing poor PA engagement and long-term health outcomes [20,21]. These limited access to PA opportunities disproportionately affect adolescents from underserved communities, who are at increased risk for deficiencies in motor proficiency, strength, cardiovascular endurance, and overall PA levels, factors closely linked to obesity and poor fitness [22]. Over time, these disparities contribute to the early onset of chronic health conditions such as cardiovascular disease and type 2 diabetes [23]. In contrast, wearable mHealth technologies offer a scalable, low-cost alternative that can be integrated into adolescents' daily routines, making them particularly promising for populations with limited access to conventional resources.

School-based physical education (PE) programs are a critical platform to provide accessible PA interventions [24]. Wearables integrated into school-based programs can promote autonomy and engagement even in constrained environments. Importantly, low levels of teacher burden have previously been associated with favorable perceptions of feasibility, further supporting the scalability of wearable devices (WD) applicability in school settings [25]. Prior studies suggest that wearables may positively influence PA levels by enhancing self-efficacy and behavioral regulation among adults [26]. Additionally, accuracy and behavioral impact of WD in general have mostly been examined in the adult population [26,27]. Despite the growing popularity of consumer wearables, little is known about long-term adherence to wearable technologies among high school age adolescents from underserved racial and ethnic minority groups [11,28]. There is a significant gap in understanding adherence behavioral patterns among adolescents facing environmental barriers, such as unsafe neighborhoods, low PA engagement, and limited access to recreational spaces.

Recent shifts in PA research emphasize the importance of broader contextual factors, both environmental (ie, social environments, physical settings, and access to resources) and personal (ie, enjoyment, autonomy, and perceived competence), in shaping PA behavior [29]. For example, previous studies found that wearables may be less accessible or sustainable for adults in underserved settings owing to socioeconomic barriers, limited internet access, and concerns about privacy and crime safety [30]. Additionally, research on long-term adherence remains limited, particularly among marginalized youth [31,32]. These contextual challenges may influence how adolescents engage with digital health tools, and the findings from broader populations may not be generalizable to these groups.

Bandura's Social Cognitive Theory (SCT) provides a behavioral foundation for understanding adherence to WD use, emphasizing how personal factors (self-efficacy and motivation), behavioral patterns (consistent device use), and environmental influences (social support and accessibility) interact to influence health behaviors [33]. The International Classification of Functioning, Disability, and Health (adherence) complements this perspective by situating these behaviors within a biopsychosocial framework, capturing how an individual's body functions and structures, activities and participation, and contextual factors (personal and environmental) collectively shape overall functioning and engagement in daily life [34]. Together, SCT and the International Classification of Functioning and Disability (ICF) informed the selection of study variables, including self-efficacy, readiness for change, perceived neighborhood safety, and physical fitness and function, allowing for a nuanced exploration of WD adherence and PA engagement among underserved adolescents.

To address the gap in understanding the influence of contextual factors on WD adherence, this study investigated associations among contextual factors, including sociodemographic characteristics, environmental conditions, PA behavior, and adherence to WD, among a racially and socioeconomically diverse cohort of high school students enrolled in a school-based PE program. Guided by the ICF framework, we included predictors, such as self-efficacy, readiness for change, perceived neighborhood safety, and physical fitness, each representing key domains within the ICF [34]. These factors have been validated in prior studies and are particularly relevant to understanding adolescent engagement in PA as they reflect the complex interplay between individual capabilities, motivation, and contextual influences [35]. Specifically, we examine the following: 1. whether demographic, personal, and environmental factors influenced WD adherence; and 2. Whether WD adherence was associated with differences in self-reported PA, fitness, and functional performance outcomes. Our findings aim to inform the design of future school-based wearable interventions that effectively engage adolescents from historically marginalized populations and promote PA through sustained behavioral changes.

Methods

Overview

This prospective cohort study was conducted in collaboration with local nonprofit organizations and public-school administrators. Participants were recruited using a purposive sampling strategy designed to ensure the representation of racially and socioeconomically diverse adolescents in an underserved community. School-based physical therapists and PE teachers facilitated recruitment through in-class announcements and take-home informational flyers distributed during sessions of state-mandated Health Opportunities Through Physical Education (HOPE) courses. This course emphasizes lifelong fitness, healthy behavior, and personal responsibility [36]. All students enrolled in the HOPE were invited to participate, minimizing selection bias by including students regardless of their prior PA engagement or academic standing. To reduce sampling bias and ensure equitable participation, parent consent and student assent materials emphasized the voluntary nature of the study, clarified that participation would not affect academic standing, and explained the privacy protections in place. Materials were designed to be culturally appropriate and accessible, with bilingual staff available to address questions for parents or guardians. These steps were taken to minimize the likelihood that guardians would restrict participation based on personal beliefs, misconceptions about wearable technology, or concerns about data privacy.

Although the school served as the central hub for recruitment and data collection, participants wore WD during and outside school hours, enabling continuous monitoring of PA across multiple contexts, including home and community settings. Predictors of adherence, such as neighborhood safety, team sports participation, and self-reported PA behaviors, reflect influences beyond the school setting. Nevertheless, the school played a pivotal role in supporting this study through structured integration of the PE curriculum, device distribution, and ongoing student engagement.

Study Timeline

Data collection spanned the entire academic year, with baseline assessments conducted in the first quarter and follow-up assessments conducted in the final quarter. This extended duration enabled evaluation of adherence to WD use throughout the academic year.

Procedures

Baseline assessments were conducted over 2 PE class sessions. Day 1 included demographic and PA surveys, and day 2 involved physical performance testing. In collaboration with PE instructors and school-based physical therapists, students received individualized step-count goals, starting with a 10% weekly increase from their baseline average. PE teachers provided weekly reminders to wear and synchronize their Fitbit devices. Measures such as self-efficacy and readiness for change were selected based on their established predictive value in adolescent PA engagement, as supported by Bandura's SCT [33] and integration into the WHO-ICF framework. Environmental variables were included to reflect contextual influences consistent with ICF's socio-ecological approach. Table 1 provides a summary of variables assessed and associated theoretical construct.

Table . Summary of predictors, data sources, and theoretical frameworks.

Variables	Outcome measures	Theoretical framework
Demographics	• Survey during PE[a] class	• Personal factors (ICF)[b]
Self-report-physical activity level	• PACE+[c] survey [37,38] • Fitbit Flex [39,40]	• Activities and participation (ICF) • Behavioral capability and self-regulation (SCT)[d]
Physical activity level	• Fitbit Flex [39,40]	• Activities and participation (ICF) • Self-monitoring and reinforcement (SCT)
Wearable device adherence	• Fitbit Flex via Fitabase [39,40]	• Activities and participation (ICF) • Self-regulation reinforcement (SCT)
Self-efficacy and readiness for change	• PACE+ Survey [37,38,41].	• Personal factors (ICF) • Self-efficacy and expectations (SCT)
Social support and enjoyment	• PACE+ Survey [42]	• Environmental factors (ICF) • Social support and reinforcement (SCT)
MVPA frequency-self-report	• PACE+ Survey [42]	• Activities and participation (ICF) • Behavioral capability and self-regulation (SCT)
Team sports participation	• Youth Risk Behavior Survey [43-45]	• Activities and participation (ICF) • Observational learning and social support (SCT)
Neighborhood walkability	• NEWS-Y Scale [46]	• Environmental factors (ICF)
Anthropometrics	• Height, weight, and waist circumference [47,48]	• Body functions and structures (ICF)
Vital signs	• Heart rate and blood pressure [49]	• Body functions and structures (ICF)
Cardiovascular endurance	• 6-minute walk test [50,51].	• Body Functions and Activities (ICF)
Functional mobility	• Timed up and down stairs	• Activities and participation (ICF)
Balance	• Single-leg stance test [52]	• Body functions and activities (ICF)
Upper-body strength	• 90° push-ups [53].	• Body functions and activities (ICF)

[a]PE: physical education.

[b]ICF: International Classification of Functioning and Disability.

[c]PACE: Pollution Abatement Costs and Expenditures.

[d]SCT: Social Cognitive Theory.

WD Adherence

Participants received a Fitbit Flex activity tracker and information on device use, care, and data synchronization from school-based physical therapists. These devices contain accelerometers and have demonstrated moderate to excellent validity compared to research-grade accelerometers in youth [39,40]. The Fitbit data were uploaded to Fitabase, a research platform for remote device monitoring. The weekly averages of daily steps were captured using Fitbit Flex devices [54]. Adherence days were defined as days with at least 10 hours of HR monitoring data [55], and Adherence week was defined as a week with at least 3 valid days of WD use [41]. A 21-day period of WD use stratified participants as adherent or nonadherent WD users. This threshold was chosen based on prior studies that defined high adherence to WD use as achieving at least 70% of the recommended daily use over a 30 day period [56] and systematic review benchmarks of 10 hours per day criteria for high adherence to daily use [41]. These timeframes were selected to balance early feasibility with theoretical relevance, particularly in adolescent populations, where motivation and environmental influences may fluctuate.

PA and Psychosocial Measures

PA behavior and psychosocial constructs were assessed using a patient-centered assessment and counseling for exercise and

nutrition (Pollution Abatement Costs and Expenditures) survey. This validated instrument includes items on decisional balance, self-efficacy, enjoyment, and social support [37,38]. It has demonstrated acceptable reliability (ICC=0.60 - 0.82) and moderate validity with accelerometer-derived MVPA (r=0.27 - 0.40) [41]. Additionally, the Pollution Abatement Costs and Expenditures Survey assessed all participants on: days per week of 60 minutes of MVPA, readiness for change, and perceived peer or family support [42-44]. Lastly, team sports participation was measured using the 2013 Youth Risk Behavior Survey, which has shown substantial reliability (κ=0.61 - 0.91) [45].

Neighborhood Walkability

Perceived neighborhood environment was assessed using the Neighborhood Environment Walkability Scale for Youth (NEWS-Y) [46]. The subscales measure safety from crime, traffic hazards, and access to land. Items were rated on a 4-point Likert scale, with higher scores indicating lower walkability. The NEWS-Y has demonstrated good test-retest reliability (ICC=0.56 - 0.87) and validity, as evidenced by associations with walking behaviors among youth [46].

Clinical Assessments

To evaluate physical fitness and functional performance, a series of standardized assessments was administered during PE class sessions by trained physical therapists. These measures were selected based on their reliability, validity, and relevance to the adolescent population.

Anthropometrics and Vital Signs

Height, weight, and waist circumference data were collected using standardized procedures [47]. BMI was calculated using CDC growth charts. The waist-to-height ratio (WtHR) was calculated; a ratio ≥0.5 indicated elevated central adiposity [48]. Resting HR and blood pressure were measured using standard clinical procedures and equipment [49].

Functional Fitness Assessments

Cardiovascular endurance was assessed using the 6-minute walk test, a submaximal exercise test that measures the distance an individual can walk in 6 minutes. The 6-minute walk test is widely used in pediatric populations and has demonstrated excellent reliability (ICC=0.94) and strong validity for estimating aerobic capacity [50,51].

Functional mobility was assessed using the timed up and go test, which evaluates lower-extremity strength, coordination, and mobility. Participants were timed as they ascended and descended stairs. The timed up and down stairs test has shown excellent reliability (ICC=0.978 - 0.999) and correlates well with other mobility and balance measures [57].

Balance was assessed using the timed single-leg stance test, in which participants stood on one leg for as long as possible without support. This test provides insight into postural control and neuromuscular stability and demonstrates acceptable reliability in youth populations [52].

Upper-body strength and endurance were assessed using 90° push-ups, in which participants performed push-ups until exhaustion with proper form. This test is a reliable indicator of muscular endurance (ICC=0.93) and has demonstrated moderate validity when compared to bench press performance (r=0.64; males and r=0.28; females) [53].

Data Analysis

Statistical analyses were performed using SPSS v28. Descriptive statistics were used to characterize the participants. Chi-square and independent t tests were used to compare the adherence groups in terms of demographic and outcome variables. Pearson correlation coefficient was used to assess the association between PA behaviors and psychosocial variables. Relationships were classified by effect size: 0.00 - 0.25 (none), 0.25 - 0.50 (weak), 0.50 - 0.75 (moderate), and >0.75 (strong) [58]. Binomial logistic regression was used to examine the associations between WD adherence and demographic, behavioral, and environmental predictors, including PA level, obesity status, perceived neighborhood safety, and nativity. A priori power analysis using G*Power (v3.1.9.6) determined a minimum sample size of 57 to detect medium effect sizes (Cohen d=0.5, f^2=0.30) with α=.05 and 80% power.

Ethical Considerations

This study was approved by the Western Institutional Review Board and the Research Ethics Committee of the County School Board. Informed consent was obtained from all the participants and their guardians. The eligibility criteria were high school adolescents aged 14 - 18 years who were enrolled in a HOPE course. Baseline assessments were conducted in the first academic quarter, and follow-up assessments were conducted in the final quarter. Surveys and physical fitness tests were conducted during the PE class. Race, ethnicity, and nativity (US or foreign-born) were self-reported. All collected data were stored on secure, password-protected servers; identifiable information was removed prior to analysis to ensure participant privacy and confidentiality. No financial or academic compensation was provided, and participation had no effect on course grades.

Results

Cohort Characteristics

Participant characteristics by WD adherence (N=63) are presented in Table 2. Of the total sample, 73% (46/63) met the WD adherence criteria. The average age was 14.8 (SD 1.17) years, with the majority in ninth grade (51/63, 81%). More than half of the participants were female (35/63, 56%), and 73% (46/63) received free or reduced lunch. Most were Black or African Americans (40/63, 72%), followed by Hispanics (16/63, 42%). Sixteen percent (10/63) were foreign-born, and 37% (23/63) reported maternal nativity outside the United States. The average BMI and WtHR for the sample were 23.06 (SD 4.62) kg/m² and 0.45 (SD 0.07), respectively.

Table . Demographic characteristics by wearable adherence, using wearable device (WD) and not using wearable device (N-WD).

Demographic characteristics	Total (N=63)	WD (n=46)	N-WD (n=17)	*P* value[a]
Age (years), mean (SD)	14.79 (1.17)	14.67 (1.06)	15.11 (1.41)	.13
Sex, n (%)				.41
Male	28 (44.4)	19 (41)	9 (53)	
Female	35 (55.6)	27 (59)	8 (47)	
Grade level, n (%)				.20
9th	51 (81)	40 (87)	11 (65)	
10th	2 (3.2)	1 (2.2)	1 (5.9)	
11th	3 (4.8)	1 (2.2)	2 (11.11)	
12th	7 (11.1)	4 (8.9)	3 (16.7)	
Socioeconomics, n (%)				
Free or reduced lunch	46 (73)	—[b]	—	—
Race and ethnicity, n (%)				
Black or African American	40 (73)	28 (72)	12 (75)	.55
Hispanic	16 (42)	14 (47)	2 (25)	.25
Student birth country (non-US)	10 (16)	9 (21)	1 (7)	.24
Parents birth country (1 - 2, non-US)	26 (41)	22 (48)	4 (24)	.07
Maternal birth country, n (%)				—
United States	30 (48)	19 (41)	11 (65)	
Bahamas	1 (1.6)	1 (2.2)	—	
Brazil	3 (4.8)	3 (6.5)	—	
Dominican Republic	1 (1.6)	1 (2)	—	
Ecuador	1 (1.6)	0	1 (7)	
Germany	1 (1.6)	1 (2)	—	
Guatemala	1 (1.6)	1 (2)	—	
Haiti	8 (13)	7 (15)	1 (7)	
Honduras	1 (1.6)	1 (2)	—	
Jamaica	2 (3)	2 (4)	—	
Mexico	3 (4.8)	2 (4)	1 (7)	
Peru	1 (1.6)	1 (2)	—	
Country not reported	10 (16)	7 (15)	3 (18)	

[a]Chi-square *P* value of ≤.05 considered statistically significant

[b]Not applicable.

Anthropometric Characteristics

As detailed in Table 3, 19% of participants were classified as overweight (12/63) and 15.9% were classified as obese (10/63). Also, 21% (13/63) had a WtHR >0.5, indicating elevated cardiometabolic risk. Significant differences between the WD adherence groups were found in resting HR (76.4 vs 67.4 bpm; *P*=.009) and single-leg stance duration (57 vs 59 s; *P*=.03). No statistically significant differences were found in BMI, WtHR, blood pressure, 6 MWT, timed up and down stairs, push-ups, or sit-to-stand performance.

Table . Health outcomes characteristics by wearable adherence status, using wearable device (WD) and not using wearable device (N-WD).

Characteristics	Total (N=63)	WD (n=46)	N-WD (n=17)	*P* value[a]
Anthropometrics				
BMI (kg/m^2), mean (SD)	23.06 (4.62)	23.16 (4.9)	22.79 (3.9)	.78
BMI-for-age (percentile), mean (SD)	66.25 (9.02)	66.1 (30.7)	66.64 (24.6)	.96
Underweight (<5 percentile), n (%)	3 (4.8)	3 (6.5)	0 (0)	.38
Healthy weight (5-<85 percentile), n (%)	36 (57)	25 (54)	11 (65)	.38
Overweight (≧85 percentile but ≦95 percentile), n (%)	12 (19)	8 (17)	4 (24)	.38
Obese (≧95 percentile), n (%)	10 (15.9)	9 (20)	1 (6)	.38
WHtR[b], mean (SD)	0.45 (0.07)	0.46 (0.08)	0.44 (0.04)	.20
≦0.5 normal risk, n (%)	49 (77.8)	34 (74)	15 (88)	.09
>0.5 higher risk, n (%)	13 (21)	12 (26)	1 (6)	.09
Vitals, mean (SD)				
Resting HR[c] (beats/min)	73.95 (14.3)	76.4 (15.3)	67.4 (9.1)	.01
SBP[d] (mm Hg)	114.7 (14.4)	115.2 (14.7)	113.4 (14.1)	.68
DBP[e] (mm Hg)	73.34 (12.9)	74.2 (13.6)	71.1 (10.9)	.42
Performance health outcomes, mean (SD)				
6MWT[f] (m)	503.14 (62.72)	496.5 (66.1)	520.6 (50.6)	.19
SLS[g] (right+left, 60s max)	57.41 (6.21)	56.9 (6.9)	59.4 (1.4)	.03
Timed up and down stairs (s)	9.27 (3.68)	9.6 (4.1)	8.3 (1.5)	.23
Push-ups repetitions	13.77 (10.36)	12.6 (10.5)	18 (8.8)	.11
Sit-to-stand repetitions per minute	36.7 (7.36)	36.13 (7.6)	38.47 (6.52)	.29

[a]*P* values (*P*<.05) were calculated on the basis of *t* test, *P* value of ≤.05 considered statistically significant.

[b]WHtR: waist to height ratio.

[c]HR: heart rate.

[d]SBP: systolic blood pressure.

[e]DBP: diastolic blood pressure.

[f]6MWT: six-minute walk test.

[g]SLS: single-leg stance.

PA and Environmental Perceptions

Table 4 summarizes the psychosocial and environmental characteristics based on WD adherence. On average, participants reported engaging in ≥60 minutes of MVPA at 2.86 (SD 2.17) days per week. Only 13% (8/63) participants met the daily MVPA guidelines, with males being more likely than females to do so (7/28, 25% vs 2/35, 5.7%; *P*=.06). Notably, 24% (15/63) of the participants reported no MVPA during the week, with inactivity being more prevalent among female adolescents (13/35, 36% vs 3/28, 12%). Participation in team sports was positively associated with MVPA; among the 33 adolescents who participated in sports, only 3 reported no MVPA. However, WD-adherent participants reported fewer days of MVPA (2 d vs 4 d; *P*=.004), fewer days of muscle-strengthening activities (3 vs 4.8 d; *P*=.004), and lower team sports participation (21/46, 46% vs 12/17, 71%; *P*=.004) compared to their nonadherent peers. No significant differences were found in psychosocial measures such as self-efficacy, family or friend support, or decisional balance. Interestingly, adherent participants perceived their neighborhoods as safer, with lower scores for crime (1.7

vs 2.4; *P*= .02), traffic concerns (2.1 vs 2.4; *P*=.04), and reported greater access to land-use features (3.1 vs 2.9; *P*=.04).

Table . Physical activity–related psychosocial characteristics and perceptions of neighborhood crime and walkability by wearable adherence status, using wearable device and not using wearable device.

Characteristics	Total (N=63)	WD[a] (n=46)	N-WD[b] (n=17)	*P*[c] value
PA[d] steps per day				
PA wearables, average steps/day, mean (SD)	—[e]	7875 (2696.23)	—	—
PA<10,000 steps/day, n (%)	—	38 (83)	—	—
PA <7000 steps/day, n (%)	—	19 (41)	—	—
PA <5000 steps/day, n (%)	—	7 (15)	—	—
PA self-report measures				
PA 60 min-MVPA[f] days<*7* days, n (%)[g]	54 (86)	42 (91)	12 (71)	.04
PA 60 min-MVPA days<*3* days, n (%)[g]	29 (46)	26 (57)	3 (18)	.006
PA 60 min-MVPA days 0 days, n (%)	15 (24)	13 (28)	2 (12)	.17
PA 60 min-MVPA days, mean (SD)	2.86 (2.17)	2.39 (2.07)	4.12 (1.97)	.004
PA muscle strengthening days, mean (SD)	3.49 (2.29)	3 (2.02)	4.82 (2.48)	.004
PA team sports participation, n (%)[g]	33 (52)	21 (46)	12 (71)	.08
PA team sports (number of teams), mean (SD)	1.94 (1.11)	1.76 (1.02)	2.41 (1.22)	.02
PA stage of change, mean (SD)	2.89 (1.01)	2.78 (.97)	3.23 (1.10)	.06
PA con, mean (SD)	1.81 (.75)	1.82 (.77)	1.76 (.71)	.40
PA pro, mean (SD)	3.64 (1.01)	3.77 (1)	3.29 (.97)	.06
PA self-efficacy, mean (SD)	3.12 (1.06)	3.04 (1.12)	3.37 (.82)	.15
PA-family support, mean (SD)	2.68 (1.34)	2.57 (1.12)	3 (1.42)	.12
PA-friend support, mean (SD)	2.95 (1.05)	2.90 (0.91)	3.09 (0.86)	.48
Neighborhood Walkability Scale, mean (SD)				
Land-use mix–diversity[h]	3.77 (0.83)	3.80 (0.79)	3.69 (0.96)	.65
Neighborhood recreation facilities[h]	3.61 (0.95)	3.59 (0.94)	3.66 (1)	.80
Residential density[h]	3.14 (1.10)	3.18 (1.10)	3.07 (1.14)	.74
Land-use mix–access[h]	3.04 (0.41)	3.10 (0.39)	2.86 (0.40)	.04
Street connectivity[h]	2.95 (0.61)	3 (0.58)	2.81 (0.68)	.30
Walking or cycling facilities[h]	2.99 (0.58)	2.91 (0.58)	3.20 (0.51)	.08
Neighborhood esthetics[h]	2.68 (0.58)	2.73 (0.60)	2.52 (0.51)	.22
Pedestrian and automobile traffic safety[i]	2.23 (0.41)	2.17 (0.43)	2.41 (0.31)	.04

Characteristics	Total (N=63)	WD[a] (n=46)	N-WD[b] (n=17)	*P*[c] value
Crime safety[i]	1.79 (0.68)	1.68 (0.61)	2.13 (0.78)	.02

[a]WD: wearable device.

[b]N-WD: not-using wearable device.

[c]*P* value of ≤.05 considered statistically significant.

[d]PA: physical activity.

[e]Not available.

[f]Activities require moderate physical effort and cause small increases in breathing or heart rate to vigorous physical activities that require hard physical effort and cause large increases in breathing or heart rate.

[g]Pre-not meeting Center for Disease Control and Prevention recommendations.

[h]Higher score=higher walkability.

[i]Higher score=lower walkability.

Predictors of Wearable Adherence

The results of the binomial logistic regression analysis in Table 5 indicate that the model was significant (χ^2_6=16.23; *P*=.01), explaining 61% of the variance (Nagelkerke R^2). Overall, 2 significant predictors emerged: perceived neighborhood crime safety and MVPA frequency. Adolescents engaging in <3 days per week of MVPA had significantly higher odds of WD adherence (OR 6.1, 95% CI 1.5-24) than those who met or exceeded the 3-day threshold. These findings suggest that lower levels of PA and perceptions of neighborhood safety may influence engagement with WD.

Table . Logistic regression predicting the likelihood of wearable device adherence.

Variables	β	SE	Wald (*df*)	Sig.	Exp (β)	95% CI for Exp (β)
Step 1						
Sex	−2.685	2.238	1.439 (1)	0.230	0.068	0.001-5.486
Ethnicity	1.478	1.498	0.974 (1)	0.324	4.386	0.233-82.609
Obesity status	3.036	2.178	1.943 (1)	0.163	20.813	0.291-1486.464
HR[a] (rest)	.057	0.069	0.685 (1)	0.408	1.059	0.925-1.211
PA[b] self-report	−1.234	0.620	3.957 (1)	0.047	0.291	0.086-0.982
Crime safety	−2.442	1.094	4.986 (1)	0.026	0.087	0.010-0.742
Constant	3.966	6.569	0.365	0.546	52.798	—[c]

[a]HR: heart rate.

[b]PA: physical activity.

[c]Not applicable.

Despite the early engagement, adherence to Fitbit monitoring declined sharply over time. Of the 63 students who initiated device use, 42 (67%) maintained adherence for at least 2 months. By month 3, retention dropped to 29 students (46%), and by month 4, only 21 students (33%) remained. Long-term adherence was minimal: 9 students (14%) persisted through month 5 and only 3 (5%) continued through month 7. This trajectory reflects a steep early decline followed by a gradual taper, underscoring the challenge of sustaining engagement beyond the initial months.

Discussion

Principal Findings

This study offers novel insights into WD adherence among underserved adolescents in a school-based setting, thus addressing a critical gap in the literature. While prior research has emphasized the role of personal and environmental factors in shaping PA behaviors, primarily focused on homogenous or higher-resource populations [27], few studies have explored how these influences interact with technology engagement in diverse, resource-limited populations. Notably, our findings revealed that adolescents with lower baseline PA, defined as engaging in MVPA fewer than 3 days per week and those least involved in team sports were more likely to adhere to WD use, precisely the group most in need of sustained support. Furthermore, 90% (9/10) of participants with obesity adhered to WD use, underscoring the potential of wearables to engage adolescents at elevated risk for chronic health conditions. Although counterintuitive, this pattern may reflect an opportunity to drive behavior change, wherein previously inactive youth leveraged technology as a catalyst to initiate and sustain PA. These findings align with prior research, such as the RAW-PA study [25], which demonstrated the high acceptability of wearable interventions among adolescents attending socioeconomically disadvantaged schools. To further

explore the impact of WD adherence on PA behavior, we conducted a post hoc analysis of step-count trends over the academic year. WD-adherent participants with lower baseline PA demonstrated modest but consistent increases in average daily steps, suggesting a potential behavioral change. Although further research is recommended, these results support the potential for wearables to facilitate engagement and promote sustained improvements in PA levels among adolescents, who are most in need of interventions.

Furthermore, adherence to WD use is often undefined or underreported in adolescent cohorts [59], particularly in resource-limited community-based settings. Our study addressed this gap by defining adherence and examining longitudinal patterns in a population disproportionately affected by obesity and related chronic conditions, thereby distinguishing it from short-term pilot studies. WD adherence was defined as having WD HR data for more than 10 hours per day for at least 3 days of the week for a minimum of 3 weeks, which aligns with benchmarks established in the wearable adherence literature [55,56] and indicates meaningful engagement. Furthermore, the data collection spanned the academic year, providing a flexible timeline for teachers and students to initiate WD use as needed. Although the initial WD engagement was high (46/63, 73%) following deployment, adherence declined sharply after the first month, with only 14% persisting until month 5 and 5% through month 7. Our findings in this underserved population echo more recent systematic reviews reporting mixed results for long-term adherence and call for interventions that combine technology with behavioral support and environmental modifications [60,61]. These results underscore the challenge and opportunity of maintaining engagement beyond the initial novelty period and highlight the need for tailored strategies, such as non–peer-competitive gamification and personalized goal setting.

Our study also revealed a stronger than anticipated association between perceived neighborhood safety and WD adherence. Our findings support the role of baseline PA behavior and environmental perceptions in PA adherence in previous studies, while extending this knowledge to WD use among predominantly Black or African American and Hispanic high school adolescents. Moreover, schools can leverage wearable data to personalize interventions, monitor progress, and foster peer support, thereby enhancing engagement and safety. Integrating wearables into school-based programs that address environmental constraints may offer a promising strategy to promote equitable PA engagement and reduce chronic disease risk [62,63]. Collaborations between schools, policymakers, device developers, and health care professionals are essential to ensure equitable access to supportive technologies and environments [22,64,65]. Beyond behavioral approaches, environmental modifications, such as improving walkability, access to sidewalks, and expanding green spaces, can support sustained PA.

Future school-based studies should explore adaptive strategies with WD within PE and wellness classes. For example, dynamic goal adjustment and real-time feedback help maintain motivation throughout the academic year, ensuring consistent exposure and accountability [66]. Wearables may also complement these efforts by mitigating environmental barriers, enabling self-monitoring, goal setting, and engagement in safe, controlled environments such as homes or schools. To enhance long-term engagement, future interventions should prioritize personalized goal setting over gamified peer competition. Evidence suggests that normative comparisons, such as an arbitrary selection of 10,000 steps/day or competition with peers, may yield only short-term motivation and can negatively impact autonomy [67]. Supporting adolescents in setting individualized goals and encouraging self-referenced progress may foster more sustainable PA habits and better align with their personal values and contexts [67,68]. For adolescents not engaged in team sports, additional motivational strategies, including mentorship and small-group activities, may foster social connections and reduce perceived isolation [69].

Although inconclusive, our study also found higher WD adherence rates among female participants than among their male peers, further highlighting the potential influence of safety-related factors on engagement with wearable technologies and underscoring the need for future research to explore sex-specific motivators and barriers to adherence. Furthermore, our study provided insights into the unique needs of adolescents who are first- or second-generation immigrants navigating acculturation, a group that has been shown to experience sharp declines in PA over time in the United States [70]. Despite their medically underserved status, race, ethnicity, and foreign birth did not significantly influence WD adherence in this study. These findings suggest that WD technology can be a culturally neutral and accessible tool for promoting PA across diverse populations. Further research should disaggregate data by generational status and language proficiency to better support first-generation and non–English-speaking youths [71].

Limitations

This study offers valuable insights into WD adherence among underserved adolescents; however, it had several limitations that should be addressed. The unexpectedly strong influence of perceived neighborhood safety may reflect regional or demographic differences, including potential sampling bias, and suggests that environmental safety perceptions play a more prominent role in technology engagement than previously recognized. The reliance on self-report measures introduces potential recall and social desirability biases, and the cross-sectional design limits causal inferences. The relatively small sample size and concentration of participants within a single school year (51/63, 81% in ninth grade) further constrain generalizability. This age clustering may obscure developmental differences in PA behavior and technology engagement across adolescence. Additionally, variability in school schedules, seasonal factors, and academic demands may influence adherence patterns and should be considered in future studies. Future studies should aim for broader age representations and larger multisite samples to enhance external validity and capture developmental variability.

This study highlights the potential of wearable mHealth technologies to engage underserved adolescents in PA, particularly those with lower baseline activity levels and limited access to structured sports. Our findings suggest that WD

adherence is influenced not only by individual behaviors but also by contextual factors, such as perceived neighborhood safety, an often-overlooked determinant of technology engagement among youth. The counterintuitive finding that adolescents with lower baseline PA levels were more likely to adhere to WD use warrants further investigation. Future research should explore mHealth access, usability barriers, and motivational factors for sustained WD use to improve PA. Anecdotal reports from teachers indicated that some students faced challenges with WD use, including limited Wi-Fi connectivity at home or school, forgetfulness, and declining interest over time. These barriers merit further exploration to inform the design of more accessible and engaging mHealth interventions. Despite being at a critical developmental stage for establishing lifelong health behaviors, older adolescents (aged 14 y - 19 y) remain underrepresented in mHealth research. Multischool collaborations and standardized adherence metrics are essential for scaling wearable-based interventions and strengthening cross-study comparisons.

Conclusions

This study highlights the potential of wearable mHealth technologies to engage underserved adolescents in PA, particularly those with lower baseline activity levels and limited access to structured sports. Adherence to WD use was significantly associated with contextual factors, including perceived neighborhood walkability, team sports participation, and prior PA behavior. These findings underscore the importance of designing school-based WD deployment that prioritizes autonomy-supportive strategies, such as personalized goal-setting and self-referenced progress to initiate and sustain PA opportunities. Moving beyond normative comparisons and peer-based competition may enhance sustained engagement and promote equitable health outcomes among high-risk youth.

Funding

No external financial support or grants were received from any public, commercial, or not-for-profit entities for the research, authorship, or publication of this article.

Conflicts of Interest

None declared.

References

1. Physical activity. World Health Organization. URL: https://www.who.int/initiatives/behealthy/physical-activity [accessed 2026-01-14]
2. Physical activity guidelines for school-aged children and adolescents. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/physical-activity-education/guidelines/index.html [accessed 2026-01-14]
3. Reimers AK, Knapp G, Reimers CD. Effects of exercise on the resting heart rate: a systematic review and meta-analysis of interventional studies. J Clin Med 2018 Dec 1;7(12):503. [doi: 10.3390/jcm7120503] [Medline: 30513777]
4. Lindgren M, Robertson J, Adiels M, et al. Resting heart rate in late adolescence and long term risk of cardiovascular disease in Swedish men. Int J Cardiol 2018 May 15;259:109-115. [doi: 10.1016/j.ijcard.2018.01.110] [Medline: 29579585]
5. Gu X, Bao S, Moss S, Zhang T. The associations of physical activity and motor performance with executive functioning among young minority children. Med Sci Sports Exerc 2023;55(9S):507-508. [doi: 10.1249/01.mss.0000984536.86093.be]
6. Robinson LE, Stodden DF, Barnett LM, et al. Motor competence and its effect on positive developmental trajectories of health. Sports Med 2015 Sep;45(9):1273-1284. [doi: 10.1007/s40279-015-0351-6] [Medline: 26201678]
7. The 2024 United States report card on physical activity for children and youth. : Physical Activity Alliance URL: https://paamovewithus.org/wp-content/uploads/2024/10/2024-US-Report-Card-Executive-Summary_FINAL.pdf [accessed 2026-01-27]
8. Tandon PS, Kroshus E, Olsen K, Garrett K, Qu P, McCleery J. Socioeconomic inequities in youth participation in physical activity and sports. Int J Environ Res Public Health 2021 Jun 29;18(13):6946. [doi: 10.3390/ijerph18136946] [Medline: 34209544]
9. Klos L, Eberhardt T, Nigg C, Niessner C, Wäsche H, Woll A. Perceived physical environment and active transport in adolescents: a systematic review. J Transp Health 2023 Nov;33:101689. [doi: 10.1016/j.jth.2023.101689]
10. Project Play survey: family spending on youth sports rises 46% over five years. : Project Play; 2025 URL: https://projectplay.org/news/2025/2/24/project-play-survey-family-spending-on-youth-sports-rises-46-over-five-years [accessed 2026-01-14]
11. About one-in-five Americans use a smart watch or fitness tracker. : Pew Research Center; 2020 URL: https://www.pewresearch.org/fact-tank/2020/01/09/about-one-in-five [accessed 2026-01-14]
12. McDonough DJ, Su X, Gao Z. Health wearable devices for weight and BMI reduction in individuals with overweight/obesity and chronic comorbidities: systematic review and network meta-analysis. Br J Sports Med 2021 Aug;55(16):917-925. [doi: 10.1136/bjsports-2020-103594] [Medline: 33731385]
13. Tudor-Locke C, Craig CL, Beets MW, et al. How many steps/day are enough? for children and adolescents. Int J Behav Nutr Phys Act 2011 Jul 28;8(1):78. [doi: 10.1186/1479-5868-8-78] [Medline: 21798014]

14. Mayorga-Vega D, Casado-Robles C, Viciana J, López-Fernández I. Daily step-based recommendations related to moderate-to-vigorous physical activity and sedentary behavior in adolescents. J Sports Sci Med 2019 Dec;18(4):586-595. [Medline: 31827342]
15. Paluch AE, Gabriel KP, Fulton JE, et al. Steps per day and all-cause mortality in middle-aged adults in the coronary artery risk development in young adults study. JAMA Netw Open 2021 Sep 1;4(9):e2124516. [doi: 10.1001/jamanetworkopen.2021.24516] [Medline: 34477847]
16. Creaser AV, Clemes SA, Costa S, et al. The acceptability, feasibility, and effectiveness of wearable activity trackers for increasing physical activity in children and adolescents: a systematic review. Int J Environ Res Public Health 2021 Jun 8;18(12):6211. [doi: 10.3390/ijerph18126211] [Medline: 34201248]
17. Kouvari M, Karipidou M, Tsiampalis T, et al. Digital health interventions for weight management in children and adolescents: systematic review and meta-analysis. J Med Internet Res 2022 Feb 14;24(2):e30675. [doi: 10.2196/30675] [Medline: 35156934]
18. Samdal GB, Eide GE, Barth T, Williams G, Meland E. Effective behaviour change techniques for physical activity and healthy eating in overweight and obese adults; systematic review and meta-regression analyses. Int J Behav Nutr Phys Act 2017 Mar 28;14(1):42. [doi: 10.1186/s12966-017-0494-y] [Medline: 28351367]
19. Terra LF, Rezende LD, Ferreira Silva RM, et al. Interventions on barriers to the participation of adolescents in physical activity: a systematic review. Int J Environ Res Public Health 2025 May 31;22(6):881. [doi: 10.3390/ijerph22060881] [Medline: 40566312]
20. Taverno Ross SE, Sebastião E, Hall G, Bantham A. Overcoming barriers to physical activity in underserved populations. Prog Cardiovasc Dis 2021;64:64-71. [doi: 10.1016/j.pcad.2020.11.002] [Medline: 33159937]
21. Bailey ZD, Krieger N, Agénor M, Graves J, Linos N, Bassett MT. Structural racism and health inequities in the USA: evidence and interventions. Lancet 2017 Apr 8;389(10077):1453-1463. [doi: 10.1016/S0140-6736(17)30569-X] [Medline: 28402827]
22. Nunez-Gaunaurd A, Moore JG, Roach KE, Miller TL, Kirk-Sanchez NJ. Motor proficiency, strength, endurance, and physical activity among middle school children who are healthy, overweight, and obese. Pediatr Phys Ther 2013;25(2):130-138. [doi: 10.1097/PEP.0b013e318287caa3] [Medline: 23542187]
23. Pineros-Leano M, Grafft N, Aguayo L. Childhood obesity risk factors by race and ethnicity. Obesity (Silver Spring) 2022 Aug;30(8):1670-1680. [doi: 10.1002/oby.23500] [Medline: 35894074]
24. Silva ECM, Helme ZE, Silva DRP, et al. Creating active schools: what influences continuous implementation following adoption? J Phys Act Health 2026 Jan 1;23(1):51-62. [doi: 10.1123/jpah.2025-0075] [Medline: 40983312]
25. Koorts H, Salmon J, Timperio A, et al. Translatability of a wearable technology intervention to increase adolescent physical activity: mixed methods implementation evaluation. J Med Internet Res 2020 Aug 7;22(8):e13573. [doi: 10.2196/13573] [Medline: 32763872]
26. Yen HY, Liao Y, Huang HY. Smart wearable device users' behavior is essential for physical activity improvement. Int J Behav Med 2022 Jun;29(3):278-285. [doi: 10.1007/s12529-021-10013-1] [Medline: 34363130]
27. Doherty C, Baldwin M, Keogh A, Caulfield B, Argent R. Keeping pace with wearables: a living umbrella review of systematic reviews evaluating the accuracy of consumer wearable technologies in health measurement. Sports Med 2024 Nov;54(11):2907-2926. [doi: 10.1007/s40279-024-02077-2] [Medline: 39080098]
28. Kim EH, Jenness JL, Miller AB, et al. Association of demographic and socioeconomic indicators with the use of wearable devices among children. JAMA Netw Open 2023 Mar 1;6(3):e235681. [doi: 10.1001/jamanetworkopen.2023.5681] [Medline: 36995714]
29. Vella SA, Aidman E, Teychenne M, et al. Optimising the effects of physical activity on mental health and wellbeing: a joint consensus statement from Sports Medicine Australia and the Australian Psychological Society. J Sci Med Sport 2023 Feb;26(2):132-139. [doi: 10.1016/j.jsams.2023.01.001] [Medline: 36737260]
30. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443. [doi: 10.2196/22443] [Medline: 33064083]
31. Ferguson T, Olds T, Curtis R, et al. Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. Lancet Digit Health 2022 Aug;4(8):e615-e626. [doi: 10.1016/S2589-7500(22)00111-X] [Medline: 35868813]
32. Bopp T, Vadeboncoeur J. "It makes me want to take more steps": racially and economically marginalized youth experiences with and perceptions of Fitbit Zips® in a sport-based youth development program. Sport and Health, Sport and Livelihoods 2021 Oct 22;9(2) [FREE Full text]
33. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
34. International Classification of Functioning, Disability and Health (ICF). World Health Organization. URL: https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health [accessed 2026-01-23]

35. Burton AM, Cowburn I, Thompson F, Eisenmann JC, Nicholson B, Till K. Associations between motor competence and physical activity, physical fitness and psychosocial characteristics in adolescents: a systematic review and meta-analysis. Sports Med 2023 Nov;53(11):2191-2256. [doi: 10.1007/s40279-023-01886-1] [Medline: 37542607]
36. HOPE-physical education core. CPALMS. URL: https://www.cpalms.org/PreviewCourse/Preview/4051 [accessed 2026-01-14]
37. Sallis JF, Grossman RM, Pinski RB, Patterson TL, Nader PR. The development of scales to measure social support for diet and exercise behaviors. Prev Med 1987 Nov;16(6):825-836. [doi: 10.1016/0091-7435(87)90022-3] [Medline: 3432232]
38. Prochaska JJ, Sallis JF, Long B. A physical activity screening measure for use with adolescents in primary care. Arch Pediatr Adolesc Med 2001 May;155(5):554-559. [doi: 10.1001/archpedi.155.5.554] [Medline: 11343497]
39. 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. [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
40. Jones D, Crossley K, Dascombe B, Hart HF, Kemp J. Validity and reliability of the fitbit flex. Int J Sports Phys Ther 2018 Aug;13(5):860-870. [doi: 10.26603/ijspt20180860] [Medline: 30276018]
41. Chan A, Chan D, Lee H, Ng CC, Yeo AHL. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: A systematic review. Int J Med Inform 2022 Apr;160:104696. [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
42. Luszczynska A, Piko B, Januszewicz A. Self-efficacy and adolescents' health. In: Levesque RJR, editor. Encyclopedia of Adolescence: Springer International Publishing; 2018:3386-3395. [doi: 10.1007/978-3-319-33228-4_190]
43. Norman GJ, Sallis JF, Gaskins R. Comparability and reliability of paper- and computer-based measures of psychosocial constructs for adolescent physical activity and sedentary behaviors. Res Q Exerc Sport 2005 Sep;76(3):315-323. [doi: 10.1080/02701367.2005.10599302] [Medline: 16270708]
44. Carlson JA, Sallis JF, Wagner N, et al. Brief physical activity-related psychosocial measures: reliability and construct validity. J Phys Act Health 2012 Nov;9(8):1178-1186. [doi: 10.1123/jpah.9.8.1178] [Medline: 22207589]
45. Youth Risk Behavior Surveillance System (YRBSS). US Centers for Disease Control and Prevention (CDC). URL: http://www.cdc.gov/HealthyYouth/yrbs/index.htm [accessed 2026-01-14]
46. Rosenberg D, Ding D, Sallis JF, et al. Neighborhood Environment Walkability Scale for Youth (NEWS-Y): reliability and relationship with physical activity. Prev Med 2009;49(2-3):213-218. [doi: 10.1016/j.ypmed.2009.07.011] [Medline: 19632263]
47. NHANES Anthropometry Procedures Manual. : US Centers for Disease Control and Prevention; 2021 URL: https://wwwn.cdc.gov/nchs/data/nhanes/public/2021/manuals/2021-Anthropometry-Procedures-Manual-508.pdf [accessed 2026-01-27]
48. Gibson S, Ashwell M. A simple cut-off for waist-to-height ratio (0·5) can act as an indicator for cardiometabolic risk: recent data from adults in the health survey for England. Br J Nutr 2020 Mar 28;123(6):681-690. [doi: 10.1017/S0007114519003301] [Medline: 31840619]
49. Vos T, Lim SS, Abbafati C, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet 2020 Oct;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9]
50. Geiger R, Strasak A, Treml B, et al. Six-minute walk test in children and adolescents. J Pediatr 2007 Apr;150(4):395-399. [doi: 10.1016/j.jpeds.2006.12.052] [Medline: 17382117]
51. Li AM, Yin J, Yu CCW, et al. The six-minute walk test in healthy children: reliability and validity. Eur Respir J 2005 Jun;25(6):1057-1060. [doi: 10.1183/09031936.05.00134904] [Medline: 15929962]
52. Springer BA, Marin R, Cyhan T, Roberts H, Gill NW. Normative values for the unipedal stance test with eyes open and closed. J Geriatr Phys Ther 2007;30(1):8-15. [doi: 10.1519/00139143-200704000-00003] [Medline: 19839175]
53. The Cooper Institute. Fitnessgram and Activitygram Test Administration Manual, 4th edition: Human Kinetics; 2010.
54. Pixel and Fitbit keep you moving. Google Store. URL: https://www.fitbit.com/global/us/produ [accessed 2025-04-09]
55. Claudel SE, Tamura K, Troendle J, et al. Comparing methods to identify wear-time intervals for physical activity with the fitbit charge 2. J Aging Phys Act 2021 Jun 1;29(3):529-535. [doi: 10.1123/japa.2020-0059] [Medline: 33326935]
56. Paolillo EW, Lee SY, VandeBunte A, et al. Wearable use in an observational study among older adults: adherence, feasibility, and effects of clinicodemographic factors. Front Digit Health 2022;4:884208. [doi: 10.3389/fdgth.2022.884208] [Medline: 35754462]
57. Zaino CA, Marchese VG, Westcott SL. Timed up and down stairs test: preliminary reliability and validity of a new measure of functional mobility. Pediatr Phys Ther 2004;16(2):90-98. [doi: 10.1097/01.PEP.0000127564.08922.6A] [Medline: 17057533]
58. Portney LG, Watkins MP. Foundations of Clinical Research: Applications to Practice, 3rd edition: Pearson Prentice Hall; 2009.
59. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560. [doi: 10.1197/jamia.M2888] [Medline: 19390112]
60. Zhang W, Xiong K, Zhu C, Evans R, Zhou L, Podrini C. Promoting child and adolescent health through wearable technology: a systematic review. Digit Health 2024;10. [doi: 10.1177/20552076241260507] [Medline: 38868368]

61. Chen X, Wang F, Zhang H, Lin Y, Zhu S, Yang Y. Effectiveness of wearable activity trackers on physical activity among adolescents in school-based settings: a systematic review and meta-analysis. BMC Public Health 2025 Mar 18;25(1):1050. [doi: 10.1186/s12889-025-22170-z] [Medline: 40102761]
62. Plotnikoff RC, Gebel K, Lubans DR. Self-efficacy, physical activity, and sedentary behavior in adolescent girls: testing mediating effects of the perceived school and home environment. J Phys Act Health 2014 Nov;11(8):1579-1586. [doi: 10.1123/jpah.2012-0414] [Medline: 24733181]
63. van Sluijs EMF, Ekelund U, Crochemore-Silva I, et al. Physical activity behaviours in adolescence: current evidence and opportunities for intervention. Lancet 2021 Jul 31;398(10298):429-442. [doi: 10.1016/S0140-6736(21)01259-9] [Medline: 34302767]
64. Pope M. Preventing weight gain in children who are school age and African-American. Pediatr Phys Ther 2016;28(2):207-216. [doi: 10.1097/PEP.0000000000000243] [Medline: 26914717]
65. Bezner JR. Promoting health and wellness: implications for physical therapist practice. Phys Ther 2015 Oct;95(10):1433-1444. [doi: 10.2522/ptj.20140271] [Medline: 25908523]
66. Creaser AV, Frazer MT, Costa S, Bingham DD, Clemes SA. The use of wearable activity trackers in schools to promote child and adolescent physical activity: a descriptive content analysis of school staff's perspectives. Int J Environ Res Public Health 2022 Oct 28;19(21):21. [doi: 10.3390/ijerph192114067] [Medline: 36360944]
67. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on fitbits with adolescents. Am J Health Educ 2017 Sep 3;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
68. Danković G, Stantić T, Herodek R, et al. Effects of commercially available wearable devices on physical activity promotion and health in children and adolescents: systematic review. Appl Sci (Basel) 2023;13(12):7194. [doi: 10.3390/app13127194]
69. Hartman SJ, Chen R, Tam RM, Narayan HK, Natarajan L, Liu L. Fitbit use and activity levels from intervention to 2 years after: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2022 Jun 30;10(6):e37086. [doi: 10.2196/37086] [Medline: 35771607]
70. Allen ML, Elliott MN, Morales LS, Diamant AL, Hambarsoomian K, Schuster MA. Adolescent participation in preventive health behaviors, physical activity, and nutrition: differences across immigrant generations for Asians and Latinos compared with Whites. Am J Public Health 2007 Feb;97(2):337-343. [doi: 10.2105/AJPH.2005.076810] [Medline: 17138919]
71. Taverno Ross SE, Larson N, Graham DJ, Neumark-Sztainer D. Longitudinal changes in physical activity and sedentary behavior from adolescence to adulthood: comparing U.S.-born and foreign-born populations. J Phys Act Health 2014 Mar;11(3):519-527. [doi: 10.1123/jpah.2011-0359] [Medline: 23416986]

Abbreviations

HOPE: Health Opportunities through Physical Education
HR: heart rate
ICF: International Classification of Functioning and Disability
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PE: physical education
SCT: Social Cognitive Theory
WD: wearable device
WtHR: waist-to-height ratio

Edited by G Badicu; submitted 10.Jul.2025; peer-reviewed by M Fraser, P Gorp; revised version received 30.Oct.2025; accepted 16.Nov.2025; published 29.Jan.2026.

Please cite as:
Nunez-Gaunaurd A, Raya M
Assessing Wearable mHealth Adherence in Underserved Adolescents and its Associations With Physical Activity, Sports, and Safety Perceptions: Prospective Cohort Study
JMIR Mhealth Uhealth 2026;14:e80465
URL: https://mhealth.jmir.org/2026/1/e80465
doi:10.2196/80465

Efficacy of a Digital Peer Support Program on Weight Management and Mental Health in University Students With Preobesity: Randomized Controlled Trial

Xingyu Liu[1], PhD; Ting Liu[2], PhD; Tao Chen[1*], PhD; Ruisi Ma[3,4*], PhD

[1]Badminton Technical and Tactical Analysis and Diagnostic Laboratory, National Academy of Badminton, Guangzhou Sport University, Guangzhou, China

[2]School of Nursing, Sun Yat-sen University, Guangzhou, China

[3]School of Physical Education, Jinan University, No. 601 Huangpu Avenue West, Tianhe District, Guangzhou, Guangdong, China

[4]Key Laboratory of Guangdong Province, Guangzhou, China

[*]these authors contributed equally

Corresponding Author:
Ruisi Ma, PhD
School of Physical Education, Jinan University, No. 601 Huangpu Avenue West, Tianhe District, Guangzhou, Guangdong, China

Abstract

Background: Approximately one-third of university students are overweight or obese, and a similar proportion experience anxiety or depression. Despite the interrelated nature of weight and mental health, interventions rarely address these issues simultaneously in young adults. Digital peer support interventions have the potential to promote healthy lifestyle and mental well-being. However, evidence is limited on whether a digital peer-driven approach can concurrently improve weight management and mental health in university populations with preobesity.

Objective: This randomized controlled trial (RCT) evaluated the efficacy of a digital peer support program in concurrently improving weight management and mental health outcomes among university students with preobesity.

Methods: In a single-blind parallel group RCT, 216 students with preobesity were allocated equally among three 6-month arms, which were a peer support intervention, an active wellness control, and a waitlist control. The peer support arm began with an interactive online workshop followed by moderated WeChat (Tencent) group discussions, daily micro tasks, biweekly group challenges, and digital badges to reinforce engagement. The active control group received the same schedule and formats but focused on general wellness topics. The waitlist group completed the same assessments without any intervention during the study period. The primary outcome measured the change in BMI from baseline to 6 months. Secondary outcomes included weekly physical activity measured in metabolic equivalent of task minutes, self-esteem, loneliness, anxiety, and depression assessed at 0, 2, 4, and 6 months. Analyses used linear mixed effects models.

Results: Retention exceeded 90%. At 6 months, the peer-support group achieved a greater BMI reduction than the active control by 0.47 (95% CI −0.89 to −0.04) kg/m² and waitlist by 0.54 (95% CI −0.85 to −0.01) kg/m². Weekly metabolic equivalent of task-minutes was 129.5 higher than active control (95% CI 53.3-205.6) and 152.9 higher than waitlist (95% CI 68.4-237.4). Self-esteem increased by 1.81 points versus active control (95% CI 0.22-3.39) and 1.99 points versus waitlist (95% CI 0.21-3.76). Loneliness scores fell by 3.79 points relative to active control (95% CI −7.03 to −0.56) and by 5.02 points relative to waitlist (95% CI −8.38 to −1.66). No significant differences emerged for anxiety or depression.

Conclusions: A comprehensive digital peer-support program delivered via WeChat produced modest but clinically meaningful improvements in weight management, physical activity, self-esteem, and social connectedness among undergraduates with preobesity compared with wellness control and no intervention. These findings suggest that integrating peer support into scalable digital platforms can simultaneously address physical and psychosocial health in at-risk university populations.

Trial Registration: ClinicalTrials.gov NCT06966661; https://clinicaltrials.gov/study/NCT06966661

(JMIR Mhealth Uhealth 2026;14:e78960) doi:10.2196/78960

KEYWORDS

digital peer support; weight management; mental health; preobese; university students; randomized controlled trial

Introduction

Young adults with preobesity are defined by a BMI between 25 and 29.9 kg/m² [1]. This group represents a critical yet underaddressed population in the fight against obesity [2]. University students with preobesity have a high prevalence of excess weight, with studies reporting that roughly 20% - 30% of students fall into the overweight category, with an additional 5% - 10% already meeting obesity criteria [3,4]. Unlike individuals of normal weight, these students are already on the threshold of obesity, experiencing early weight gain–related health risks and often subtle metabolic changes [5]. Crucially, even at this preobese stage, individuals already experience adverse consequences, including heightened stress, increased depressive symptoms, and lower academic performance [6-8]. This underscores that focusing on students with preobesity is not merely a preventative measure but a response to an emerging health and psychosocial burden that distinctly affects this group.

Intervening during the preobese stage in emerging adulthood may offer a pivotal window to alter lifelong health trajectories [9]. The university years are a well-recognized turning point for weight gain and lifestyle habits, commonly known as the "freshman weight gain" phenomenon. Grounded in the socioecological model, first-year transitions expose students to interpersonal, organizational, and environmental cues that favor weight gain [10,11]. The prevention of further weight gain and the promotion of modest weight loss are widely viewed as more cost-effective and achievable than treating established obesity later on [12,13]. Overweight students generally have fewer entrenched behaviors and less severe physiological impairment than individuals with obesity, meaning early intervention can capitalize on the reversibility of risk [14]. From a public health perspective, preventing the progression from overweight to obesity avoids the need for intensive medical treatments and reduces long-term health care costs [12,15,16]. Focusing on college students with preobesity offers a critical window for high-impact preventive measures, potentially delaying cardiometabolic diseases and mental health decline before they become more difficult and expensive to manage [17,18].

Despite the clear rationale for early intervention, university students with preobesity face unique challenges that current interventions rarely address adequately [8,19]. This transitional life stage features new freedoms and psychosocial pressures. Academic stress, changing peer networks, and the shift from adolescence to adulthood can negatively affect eating, activity, and mental well-being [20]. Overweight students may struggle with body image concerns and social stigma while also encountering environments filled with unhealthy food options and sedentary habits [21]. These psychosocial stressors and environmental factors intertwine, often creating a vicious cycle where weight gain and mental health issues reinforce each other [22]. However, most existing health programs operate in isolation. Weight management initiatives in college settings tend to focus solely on diet and exercise [18], whereas campus mental health services rarely integrate lifestyle change for physical health [23]. There is a notable gap in interventions that treat physical and mental health as interconnected targets. In fact, a recent review highlighted that very few programs for university students have successfully improved both health behaviors, such as diet, physical activity, and weight-related outcomes, and mental health outcomes at the same time [8]. This lack of comprehensive approaches means students with preobesity at a crucial phase often do not receive support that addresses the full range of their needs, namely managing weight while concurrently bolstering psychological resilience and well-being. Moreover, no prior study has included both an attention-matched active control and a waitlist control arm to isolate specific peer-support benefits from general engagement effects.

Digital peer-support interventions, encompassing online group chats, moderated peer discussions, peer challenges, and digital badge systems, offer a promising and innovative solution to fill this gap for young adults with preobesity [24]. According to social cognitive theory, observing peers' successes and receiving encouragement enhances self-efficacy and outcome expectancies, thereby improving motivation, adherence, and outcomes in weight management [25]. In line with this theory, social support and positive peer influences operate through observational learning and modeling to normalize healthy behaviors and reduce the isolation or stigma an overweight student might feel, especially among university students who are highly receptive to peer influences [25,26]. By delivering such peer support via digital platforms, we can amplify its reach and convenience. Technology-based interventions are inherently adaptable, engaging students in real time, and digital platforms offer broad appeal and extensive reach across this population [27]. Through social media groups, mobile apps, or online forums, students can connect with "virtual" peers to share progress, setbacks, and encouragement [25]. Prior studies indicate that online peer communities can recreate the sense of camaraderie found in in-person support groups, providing accountability and empathy among participants with similar goals [25]. Moreover, digital peer support can be accessed on-demand and flexibly integrated into students' busy schedules, overcoming common barriers to participation such as time constraints or embarrassment in face-to-face settings [28]. Although this approach has been previously studied, existing trials seldom focus on the critical preobese life stage, compare multiple digital support modalities within the same randomized controlled trial (RCT), or include long-term follow-up and an expanded set of psychosocial outcomes.

In light of the above, this study aimed to evaluate the efficacy of a digital peer-support program on weight management and mental health in university students with preobesity. To test this, we conducted a single-blind, parallel-group, 3-arm RCT comparing a digital peer-support intervention, an attention-matched active control, and a waitlist control, with prespecified pairwise comparisons. We hypothesized that participants in the peer-support arm would demonstrate greater improvements in BMI and mental health measures than those in the active control arm, and that both intervention arms would outperform the waitlist control. By examining this novel intervention, our goal is to address an important gap in student health promotion and contribute evidence on the integration of digital peer support in weight management and mental health prevention efforts. Our intervention was grounded in social

cognitive theory and the socioecological model [29,30]. We anticipated that peer modeling within a supportive digital environment would enhance self-efficacy and motivation for healthy behaviors and provide a multilevel context for behavior change.

Methods

Study Design

We conducted a single-blind, parallel-group, superiority RCT entirely online over a 6-month intervention period to evaluate a brief digital peer-support intervention for weight management and mental health in university students with preobesity. The participants were drawn from 6 universities in Guangzhou, including Sun Yat-sen University, South China Normal University, Jinan University, South China Agricultural University, South China University of Technology, and Guangdong University of Foreign Studies. Participants were recruited between September 1 and October 31, 2024, and follow-up assessments were conducted at 2 months in November 2024, 4 months in January 2025, and at the study end on April 30, 2025. The study is reported in accordance with CONSORT (Consolidated Standards of Reporting Trials) guidelines and was registered prior to enrollment (ClinicalTrials.gov ID NCT06966661). Participants were recruited from university campuses via online advertisements, email bulletins, and social media postings. Interested individuals accessed a screening survey by clicking a link or QR code in the advertisement.

Participants and Recruitment

Eligible participants were currently enrolled university undergraduate students, aged 18 years or older, with preobese weight status (defined as BMI between 25.0 and 29.9 kg/m² based on self-reported height and weight at screening). Additional inclusion criteria were access to a stable internet connection and a device for video conferencing, to enable participation in the online program. Participants were excluded if they had a BMI in the obese range (≥30 kg/m^2) or normal range (<25 kg/m^2) at screening, had any diagnosed medical or psychiatric condition that would interfere with study participation or require alternate care (eg, uncontrolled diabetes, an eating disorder, psychosis, or acute suicidal ideation), or were already enrolled in a weight management program or receiving psychological counseling. To ensure safety, individuals with severe depression or high acute mental health risk were directed to clinical services and not enrolled. Pregnant students were also excluded due to potential weight changes unrelated to the intervention.

Randomization and Blinding

Eligible consenting participants were randomly assigned in a 1:1:1 ratio to 3 groups that were the digital peer support intervention group, the active control group receiving standard online wellness content, and the waitlist control group. Randomization was carried out using a computer-generated sequence. This sequence applied permuted blocks of varying sizes (6 and 9). The allocation list was prepared by an independent statistician who was not involved in participant recruitment or intervention delivery. To maintain allocation concealment, the randomization list was stored in a password-protected file on a secure server accessible only to the statistician. Study coordinators obtained the next assignment code via a secure messaging system only after participants completed baseline assessments and eligibility confirmation, which prevented recruiters from foreseeing upcoming assignments. Because the trial was conducted entirely online, sealed envelopes were not used. Instead, participants were assigned by the statistician through a secure web-based platform. The study used a single-blind design, and participants were informed that the trial compared 3 different online wellness programs but were not told which program was the primary intervention. Both the intervention and active control groups received online content and peer interaction for 6 months to control for expectation and exposure time; the waitlist control group did not receive any intervention activities during this period.

Facilitators and research staff who delivered the intervention could not be blinded due to the nature of the program. However, they were not involved in outcome assessments, and distinct facilitators managed each group without evaluative responsibilities. Follow-up data on BMI, self-esteem, loneliness, and other outcomes were collected via self-administered online surveys distributed by research assistants who were blind to group assignment and used anonymous study codes to track responses. These assistants did not have access to the randomization list. Data analysts remained blinded to group assignments until the primary analyses were completed.

Intervention Group

Participants in this arm began with a 15-minute live interactive workshop delivered synchronously via an online videoconferencing platform. All participants joined remotely from their own devices. A trained health educator used a shared digital whiteboard and chat functions to guide the group through evidence-based guidance on balanced eating, regular movement, stress management, and resilience building using real-time polls and interactive infographics to illustrate key points in a neutral, student-friendly language while participants added their own reflections and questions. Immediately afterward, teams of 4 to 6 students convened in a dedicated WeChat group for a 1-hour conversation combining text and voice interactions under the guidance of a trained moderator, with participants sharing brief reflections on personal challenges, exploring sustainable behavior change strategies, and agreeing on 1 or 2 small health goals for the coming week while the moderator used open-ended prompts to ensure inclusive and supportive dialogue. Over the 6-month follow-up, the WeChat group delivered daily micro tasks 7 days a week, such as minute-long nutrition quizzes, morning breathing exercises, and simple step count challenges. Each completed task earned participants a digital badge displayed in the group, and automated motivational messages reinforced progress and reminded students of their goals. Every 2 weeks, the moderator launched a friendly group challenge via the WeChat group, such as collectively walking 10,000 steps in a day to foster camaraderie and mutual accountability, and the WeChat group remained open for ongoing peer support with study staff overseeing the conversation to keep discussions respectful and on topic.

Active Control Group

Participants followed the same schedule and digital formats, including a 15-minute workshop, a 1-hour WeChat discussion, daily micro tasks, and 2-week challenges, but the content focused on general student wellness topics, such as effective study habits, time management, and healthy sleep. WeChat discussions addressed academic strategies, identification of barriers, and the setting of weekly improvement goals.

Waitlist Control Group

Participants in this arm completed all baseline and follow-up assessments on the same schedule as the other 2 groups but did not receive any workshop activity, WeChat discussions, or micro task prompts during the 6-month period. They were informed that they would gain access to the full peer-support program after the final data collection. Throughout the study, they continued their usual routines without additional contact beyond survey reminders and, at the conclusion of the follow-up, they were offered the opportunity to join the interactive workshop, the WeChat peer discussions, and to receive the series of daily micro tasks.

Procedures

After providing informed consent, each participant completed an online baseline survey capturing demographic information, self-reported height and weight, and all outcome measures. Participants were then randomized in equal proportions to 1 of the 3 arms and received secure instructions for scheduling their workshop and WeChat group times or confirmation of their waitlist status. Sessions were arranged to match participants' availability, and reminders were sent by email and text 1 to 2 days before each workshop or discussion session. Attendance and engagement were logged by facilitators, and all WeChat group activities were monitored by study staff to document participation. Completion of the daily micro-learning tasks and the awarding of digital badges were automatically recorded via backend logs in the WeChat mini-program. Each week, the data manager generated adherence reports to flag participants who completed fewer than half of the assigned tasks or who had not earned a badge in 7 consecutive days. These participants were contacted by study staff to encourage reengagement. No participants were excluded or reassigned on the basis of missed tasks. Adherence metrics were summarized descriptively and considered in exploratory analyses but did not influence eligibility or weighting in the primary available-case analyses. Follow-up assessments were delivered online at 2 months, 4 months, and 6 months after baseline using identical survey links to maintain blinding. Nonresponders received reminders at 5 and 12 days postinvitation and a final contact attempt by phone or messaging app if needed. Participants who completed each survey received a modest gift card honorarium and, after the final follow-up, the waitlist control group was provided access to the peer support program while the study was closed to further interventions.

Outcomes and Measures

The primary outcome was BMI, which was recorded from baseline to the 6-month assessment, with interim follow-up assessments at 2 months and at 4 months. Participants were instructed to weigh themselves in the morning before eating and to measure height against a flat wall, noting any deviations in the procedure.

Secondary outcomes encompassed changes in physical activity, anxiety, depression, mental well-being, perceived stress, general well-being, self-esteem, loneliness, and social connectedness. Physical activity was captured by the International Physical Activity Questionnaire, a 7-item measure recording the frequency and duration of walking, moderate-intensity, and vigorous activities over the previous 7 days [31]. Anxiety symptoms were assessed using the 7-item Generalized Anxiety Disorder (GAD) scale, which rates symptoms over the past 2 weeks from 0 (not at all) to 3 (nearly every day) for a total score of 0-21 [32]; the Chinese version demonstrates strong reliability, validity, and measurement invariance in university samples [33]. Depressive symptoms were measured by the 9-item Patient Health Questionnaire (PHQ), scored 0 - 3 per item for a total of 0-27 [34]; Chinese validation studies report high internal consistency and test-retest reliability [35]. Mental well-being was captured by the 14-item Warwick-Edinburgh Mental Well-being Scale (WEMWBS), with each item scored 1 - 5 for a total range of 14-70 [36]; the Chinese WEMWBS shows robust internal consistency and construct validity among nursing trainees and general students [37]. Perceived stress was assessed via the 10-item Perceived Stress Scale (PSS), rated 0 (never) to 4 (very often), yielding scores from 0 to 40 [38]; the Simplified Chinese version exhibits high reliability (α>.90) and a clear 2-factor structure [39]. General well-being was measured by the 5-item World Health Organization Well-Being Index (WHO-5), scored 0 (at no time) to 5 (all of the time) and converted to a 0 - 100 index [40]; the Chinese WHO-5 has demonstrated good internal consistency (α≈.85) and unidimensional fit in student and patient samples [41]. Self-esteem was measured by the 10-item Rosenberg Self-Esteem Scale on a 4-point agreement scale [42]; Chinese adaptations confirm robust factorial and criterion validity [43]. Loneliness was assessed by the 20-item University of California, Los Angeles (UCLA) Loneliness Scale Version 3, with items rated 1 (never) to 4 (often) [44]; recent validation of Chinese variants supports strong reliability and validity in adolescent and university populations [45]. Social connectedness was measured by the 8-item Social Connectedness Scale (SCS), scored 1 (strongly disagree) to 6 (strongly agree) [46]; the Chinese translation demonstrates satisfactory psychometric properties in student samples [47].

Statistical Analysis

Effect size assumptions (Cohen $d \approx 0.50$) were derived from published digital weight management trials [48]. A priori power analysis using G*Power (Heinrich-Heine-Universität Düsseldorf) for a 3-arm, 4 timepoint repeated measures design [49], with an assumed within-subject correlation of 0.5, 2-tailed α=.05, and 80% power, indicated a minimum sample of 159 participants (n=53 per arm). Allowing for up to 20% attrition, we therefore enrolled 216 participants (n=72 per arm), ensuring robust power to detect the expected effects.

Primary and secondary outcomes were measured at baseline and 2, 4, and 6 months. Analyses were performed on an

available-case basis. Participants who completed at least 1 follow-up assessment were included in the analysis. Missing data were less than 5% and under the missing at random assumption, linear mixed effects models estimated by maximum likelihood were used without additional imputation [50]. Each model specified fixed effects for study arm time and the study arm by time interaction and adjusted for the baseline value of the outcome. A random intercept for each participant accounted for within-subject correlation across repeated measures. The primary test of efficacy was the study arm by time interaction for BMI, and significance for this test was evaluated at 2-tailed P <.05. Secondary outcomes were analyzed in the same way, and false discovery rate–adjusted P values were reported to provide context for the full set of psychosocial measures. Model diagnostics included inspection of residuals for normal distribution and equal variance. All analyses were performed in R (version 4.2.2; R Foundation for Statistical Computing) with the *lme4* and *mice* packages, and results are presented alongside 95 % CIs [51].

Ethical Considerations

The study protocol was reviewed and approved by the Institutional Review Board of Guangzhou Sport University (approval number 2025LCLL-057), ensuring that the research adhered to institutional and national ethical guidelines for research involving human participants. All participants provided informed consent electronically after a one-on-one discussion with a research staff member about the study procedures and eligibility criteria. Participation was voluntary, and they were informed that they could withdraw from the study at any time without consequences. Data were collected via secure online surveys and stored on a password-protected server accessible only to the research team, and personal identifiers were removed before analysis so that only deidentified, aggregated data were used for publication to protect participant confidentiality. Participants did not receive monetary or material compensation, and digital badges and peer support interactions provided within the intervention were part of the program design and served as motivational tools rather than compensation.

Results

Of the 822 students assessed for eligibility, 216 met inclusion criteria and were randomized equally to the intervention (n=72), active control (n=72), and waitlist control (n=72) arms. Follow-up rates exceeded 90% at each of the 4 assessment points (Figure 1). Baseline demographic and clinical characteristics were well-balanced across groups: the overall mean age was 21.69 (SD 2.40) years, 48.15% (104/216) were female, and the mean BMI was 27.03 (SD 1.29) kg/m^2; no between-group differences reached statistical significance on any measure (Table 1).

Adherence to the program activities was high. Participants in the digital peer-support arm completed an average of 130.94 out of 181 of assigned daily micro tasks (72.34%, SD 12.38%) and attended a mean of 21.68 out of 24 scheduled WeChat group discussions (90.33%). In the active wellness control arm, participants completed a mean of 124.80 out of 181 daily tasks (68.94%, SD 13.11%) and attended 21.30 out of 24 group discussions (88.73%). These adherence estimates were obtained through daily check-ins conducted by research assistants within the WeChat groups, who recorded participant confirmations of task completion and discussion attendance in study logs rather than relying on automated device logs.

At 2 months, there were no significant group differences on any outcome. However, by 4 months (Table 2), the intervention arm had already achieved markedly higher weekly physical activity than both control groups: adjusted total metabolic equivalent of task (MET) - minutes were 124.10 MET greater than in the active control (95% CI 45.64-202.55; P<.001) and 147.32 MET greater than in the wait - list control (95% CI 62.51-232.13; P<.001). Self - esteem scores in the intervention arm exceeded those of the active control by 1.93 points (95% CI 0.35-3.52; P=.04) and those of the wait - list control by 2.08 points (95% CI 0.30-3.87; P=.04). Loneliness scores were also significantly lower among intervention participants, with scores 3.76 points below the active control (95% CI –6.93 to –.58; P=.03) and 5.09 points below the waitlist control (95% CI –8.40 to –1.78; P=.03), reflecting a greater reduction in perceived social isolation (Figure 2).

Figure 1. Flow diagram of participant recruitment, allocation, follow-up, and analysis in the randomized trial.

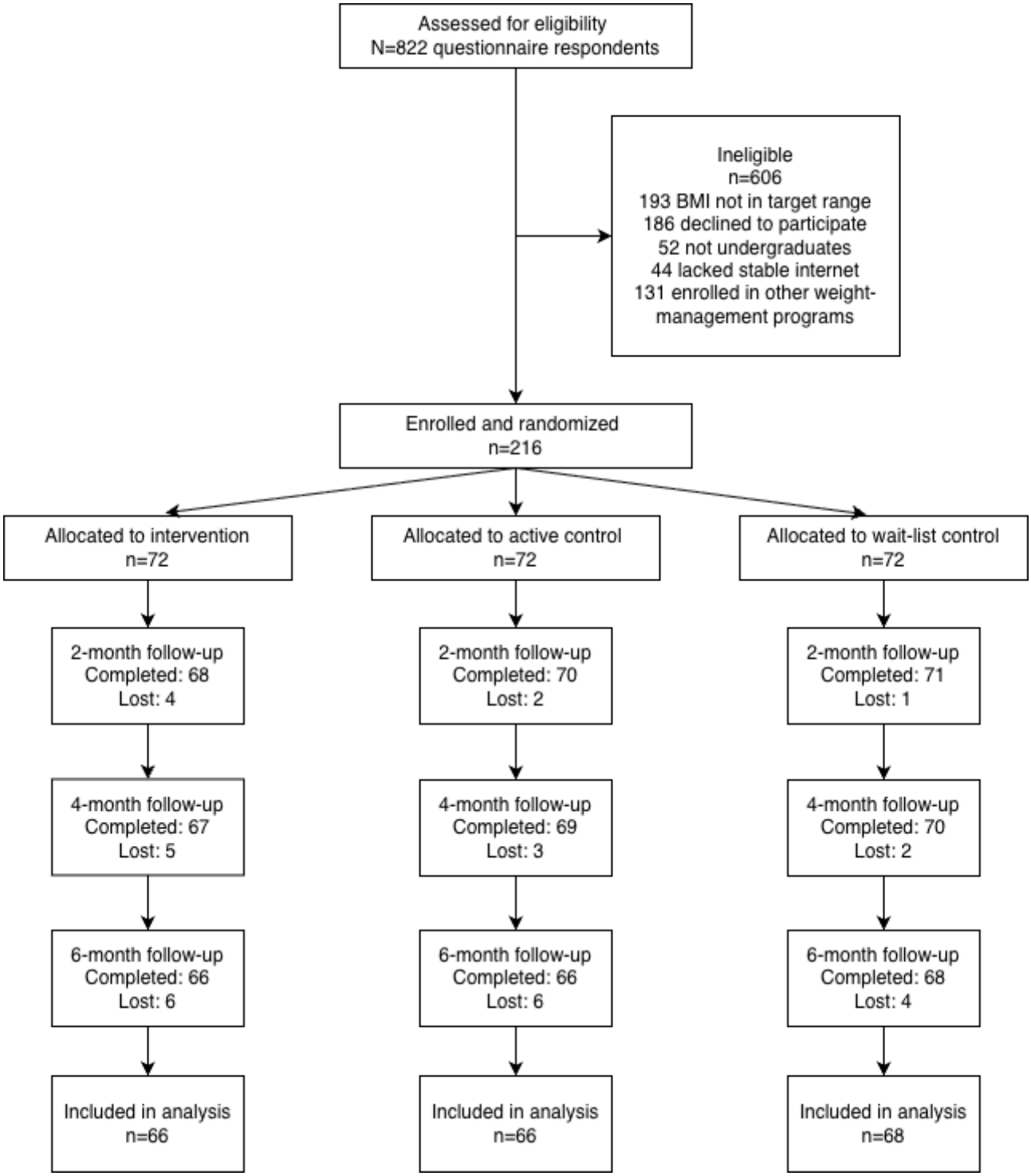


Table . Baseline demographic and psychosocial characteristics of participants, overall and by group.

Characteristics[a]	Total sample (N=216)	Intervention (n=72)	Active control (n=72)	Waitlist control (n=72)
Age, mean (SD)	21.69 (2.40)	21.83 (2.25)	21.21 (2.35)	22.04 (2.55)
Sex, n (%)				
Male	112 (51.90)	36 (50)	40 (55.56)	36 (50)
Female	104 (48.10)	36 (50)	32 (44.44)	36 (50)
Grade[b], n (%)				
Grade 1	60 (27.78)	20 (27.78)	15 (20.83)	25 (34.72)
Grade 2	44 (20.37)	15 (20.83)	16 (22.22)	13 (18.06)
Grade 3	59 (27.31)	23 (31.94)	19 (26.39)	17 (23.61)
Grade 4	53 (24.54)	14 (19.44)	22 (30.56)	17 (23.61)
WeChat[c] use (hrs/day), mean (SD)	1.96 (1.02)	2.04 (0.93)	2.03 (1.12)	1.82 (1)
BMI, kg/m^2, mean (SD)	27.03 (1.29)	26.93 (1.28)	27.10 (1.34)	27.06 (1.27)
MET[d], mean (SD)	628 (213.17)	646.40 (196.73)	612.44 (212.90)	625.14 (230.17)
GAD[e], mean (SD)	5.31 (3.62)	5.11 (3.69)	5.79 (3.78)	5.03 (3.38)
PHQ[f], mean (SD)	5.90 (4.53)	5.97 (4.53)	5.96 (5.04)	5.78 (4.05)
WEMWBS[g], mean (SD)	50.30 (9.45)	49.72 (8.77)	50.19 (9.75)	50.99 (9.88)
PSS[h], mean (SD)	16.83 (5.79)	16.97 (6.10)	17.25 (5.79)	16.28 (5.53)
WHO-5[i], mean (SD)	61.33 (20.15)	60.67 (19.77)	64.15 (19.63)	59.17 (20.99)
Self-esteem[j], mean (SD)	30.54 (4.84)	30.54 (5.13)	30.57 (4.06)	30.51 (5.30)
Loneliness[k], mean (SD)	40.68 (9.62)	41 (9.71)	40.07 (9.56)	40.97 (9.68)

[a]Values are presented as mean (SD) for continuous variables and n (%) for categorical variables.

[b]Grade denotes year of study.

[c]WeChat use reflects average daily hours on the platform.

[d]MET: metabolic equivalent task minutes per week, which was assessed via the International Physical Activity Questionnaire.

[e]GAD: Generalized Anxiety Disorder-7.

[f]PHQ: Patient Health Questionnaire-9.

[g]WEMWBS: Warwick-Edinburgh Mental Well-being Scale.

[h]PSS: Perceived Stress Scale.

[i]WHO‐5: 5-item World Health Organization Well‐Being Index.

[j]Self‐Esteem: Rosenberg Self‐Esteem Scale.

[k]Loneliness: University of California, Los Angeles Loneliness Scale.

Table . Adjusted linear mixed - effects model estimates with group× time interaction and pairwise contrasts.

Outcomes and timepoint[a]	Intervention	Active control	Waitlist control	Int (LRT)[b], χ^2 (df)	Intervention versus AC[c], mean difference (95% CI)	Intervention versus WLC[d], mean difference (95% CI)
BMI, kg/m²						
Baseline	26.93 (1.28)	27.10 (1.34)	27.06 (1.27)	51.57 (6)	–0.17 (–0.59 to 0.26)	–0.13 (–0.55 to 0.29)
2 months	26.84 (1.27)	27.06 (1.32)	27.02 (1.28)	—[e]	–0.21 (–0.63 to 0.21)	–0.18 (–0.60 to 0.24)
4 months	26.80 (1.25)	27.03 (1.37)	26.99 (1.32)	—	–0.24 (–0.66 to 0.18)	–0.19 (–0.61 to 0.23)
6 months	26.63 (1.28)	27.10 (1.34)	27.06 (1.27)	—	–0.47 (–0.89 to –0.04)	–0.54 (–0.85 to –0.01)
MET[f]						
Baseline	646.40 (196.73)	612.44 (212.90)	625.14 (230.17)	21 (6)	33.96 (−33.11 to 101.03)	21.26 (−48.80 to 91.32)
2 months	690.64 (243.11)	632.54 (228.78)	608.71 (268.26)	—	58.10 (−19.15 to 135.34)	81.93 (−1.84 to 165.70)
4 months	719.26 (247.92)	595.17 (231.28)	598.81 (196.49)	—	124.10 (45.64-202.55)	147.32 (62.51-232.13)
6 months	749.21 (205.77)	600.42 (231.22)	603.60 (196.60)	—	129.45 (53.28-205.61)	152.87 (68.35-237.38)
GAD[g]						
Baseline	5.11 (3.69)	5.79 (3.78)	5.03 (3.38)	13.04 (6)	–0.68 (–1.83 to 0.47)	0.08 (–1.07 to 1.24)
2 months	4.67 (3.63)	5.15 (3.84)	4.38 (3.40)	—	–0.49 (–1.64 to 0.67)	0.29 (–.86 to 1.45)
4 months	4.24 (3.55)	4.75 (3.73)	3.72 (3.09)	—	–0.51 (–1.67 to 0.64)	0.51 (–0.64 to 1.67)
6 months	3.79 (3.67)	4.38 (3.78)	3.06 (3.06)	—	–0.58 (–1.74 to 0.57)	0.74 (–0.42 to 1.89)
PHQ[h]						
Baseline	5.97 (4.53)	5.96 (5.04)	5.78 (4.05)	6.33 (6)	0.01 (–1.31 to 1.33)	–0.08 (–1.40 to 1.25)
2 months	5.49 (4.33)	5.40 (4.80)	5.25 (4.02)	—	–0.49 (–1.81 to 0.83)	–0.42 (–1.74 to 0.90)
4 months	4.78 (4.41)	4.85 (4.81)	4.71 (3.87)	—	–0.51 (–1.83 to 0.81)	–0.74 (–2.06 to 0.57)
6 months	4.12 (4.33)	4.46 (4.62)	3.83 (3.70)	—	–0.51 (–1.83 to 0.81)	–0.65 (–1.97 to 0.67)
WEMWBS[i]						
Baseline	49.72 (8.77)	50.19 (9.75)	50.99 (9.88)	7.21 (6)	–0.47 (–4 to 3.06)	–2.67 (–6.20 to 0.87)
2 months	49.54 (9.19)	50.50 (9.79)	50.76 (10.26)	—	–1.60 (–5.13 to 1.93)	–1.42 (–4.95 to 2.11)

Outcomes and timepoint[a]	Intervention	Active control	Waitlist control	Int (LRT)[b], χ^2 (df)	Intervention versus AC[c], mean difference (95% CI)	Intervention versus WLC[d], mean difference (95% CI)
4 months	49.08 (9.52)	50.25 (9.42)	50.62 (10.63)	—	–2.49 (–6.01 to 1.03)	–0.87 (–4.40 to 2.67)
6 months	48.60 (9.82)	50.04 (9.66)	50.57 (10.44)	—	–1.51 (–5.04 to 2.02)	0.11 (–3.42 to 3.65)
PSS[j]						
Baseline	16.97 (6.10)	17.25 (5.79)	16.28 (5.53)	7.46 (6)	–0.28 (–2.29 to 1.73)	–0.97 (–2.98 to 1.04)
2 months	16.21 (6.46)	16.93 (5.86)	15.53 (5.75)	—	–0.42 (–2.42 to 1.58)	–0.93 (–2.94 to 1.07)
4 months	15.50 (6.87)	16 (6.18)	14.99 (6.15)	—	–0.76 (–2.77 to 1.25)	–1.67 (–3.68 to 0.34)
6 months	14.54 (7.03)	15.56 (6.46)	14.42 (6.51)	—	–0.90 (–2.90 to 1.10)	–0.90 (–2.91 to 1.11)
WHO-5[k]						
Baseline	60.67 (19.77)	64.15 (19.63)	59.17 (20.99)	2.95 (6)	–3.49 (–10.42 to 3.45)	–1.78 (–8.71 to 5.15)
2 months	62.18 (20.64)	64.86 (19.77)	59.65 (22.38)	—	–1.11 (–8.04 to 5.82)	1.75 (–5.18 to 8.68)
4 months	62.54 (21.60)	64.96 (19.89)	60.21 (22.37)	—	–1.18 (–8.11 to 5.74)	3.75 (–3.18 to 10.68)
6 months	62.04 (22.67)	65.75 (19.89)	60.90 (22.47)	—	–0.31 (–7.24 to 6.62)	2.58 (–4.35 to 9.51)
Self-esteem[l]						
Baseline	30.54 (5.13)	30.57 (4.06)	30.51 (5.30)	5.60 (6)	–0.03 (–1.55 to 1.50)	0.03 (–1.69 to 1.75)
2 months	30.06 (5.27)	30.43 (4.15)	29.93 (5.31)	—	–0.38 (–1.94 to 1.19)	0.12 (–1.62 to 1.87)
4 months	31.82 (5.41)	29.89 (4.11)	29.74 (5.42)	—	1.93 (0.35-3.52)	2.08 (0.30-3.87)
6 months	31.44 (5.48)	29.64 (4.03)	29.46 (5.27)	—	1.81 (0.22-3.39)	1.99 (0.21-3.76)
Loneliness[m]						
Baseline	41 (9.71)	40.07 (9.56)	40.97 (9.68)	3.07 (6)	0.93 (–2.24 to 4.11)	0.03 (–3.17 to 3.22)
2 months	40.03 (9.93)	39.29 (9.78)	40.26 (10.07)	—	0.74 (–2.51 to 3.98)	–0.24 (–3.53 to 3.06)
4 months	34.49 (9.77)	38.25 (9.52)	39.58 (10.30)	—	–3.76 (–6.93 to -.58)	5.09 (–8.40 to –1.78)

Outcomes and timepoint[a]	Intervention	Active control	Waitlist control	Int (LRT)[b], χ^2 (df)	Intervention versus AC[c], mean difference (95% CI)	Intervention versus WLC[d], mean difference (95% CI)
6 months	33.55 (9.94)	37.35 (9.69)	38.57 (10.45)	—	–3.79 (–7.03 to –0.56)	–5.02 (–8.38 to –1.66)

[a]Adjusted means (SE) are estimated from linear mixed‐effects models with a random intercept for each participant, including fixed effects for group (1 = intervention, active control, 3=waitlist control), time (baseline, 2 months, 4 months, 6 months), and their interaction. Interaction effects were tested via likelihood‐ratio tests (chi-square (df)); reported chi-square and degrees of freedom refer to the comparison between the full model (with group × time) and the reduced model (without the interaction). Pairwise contrasts (intervention vs active control; intervention vs waitlist control) give mean differences and 2‐sided 95% CIs. Positive values indicate higher scores in the intervention group.

[b]LRT: likelihood ratio test.

[c] AC: active control.

[d] WLC: waitlist control.

[e]Not applicable.

[f]MET: total weekly metabolic equivalent minutes.

[g]GAD: Generalized Anxiety Disorder‐7.

[h] PHQ: Patient Health Questionnaire‐9.

[i]WEMWBS: Warwick–Edinburgh Mental Well‐being Scale.

[j]PSS: Perceived Stress Scale.

[k]WHO‐5: 5-item World Health Organization Well‐Being Index.

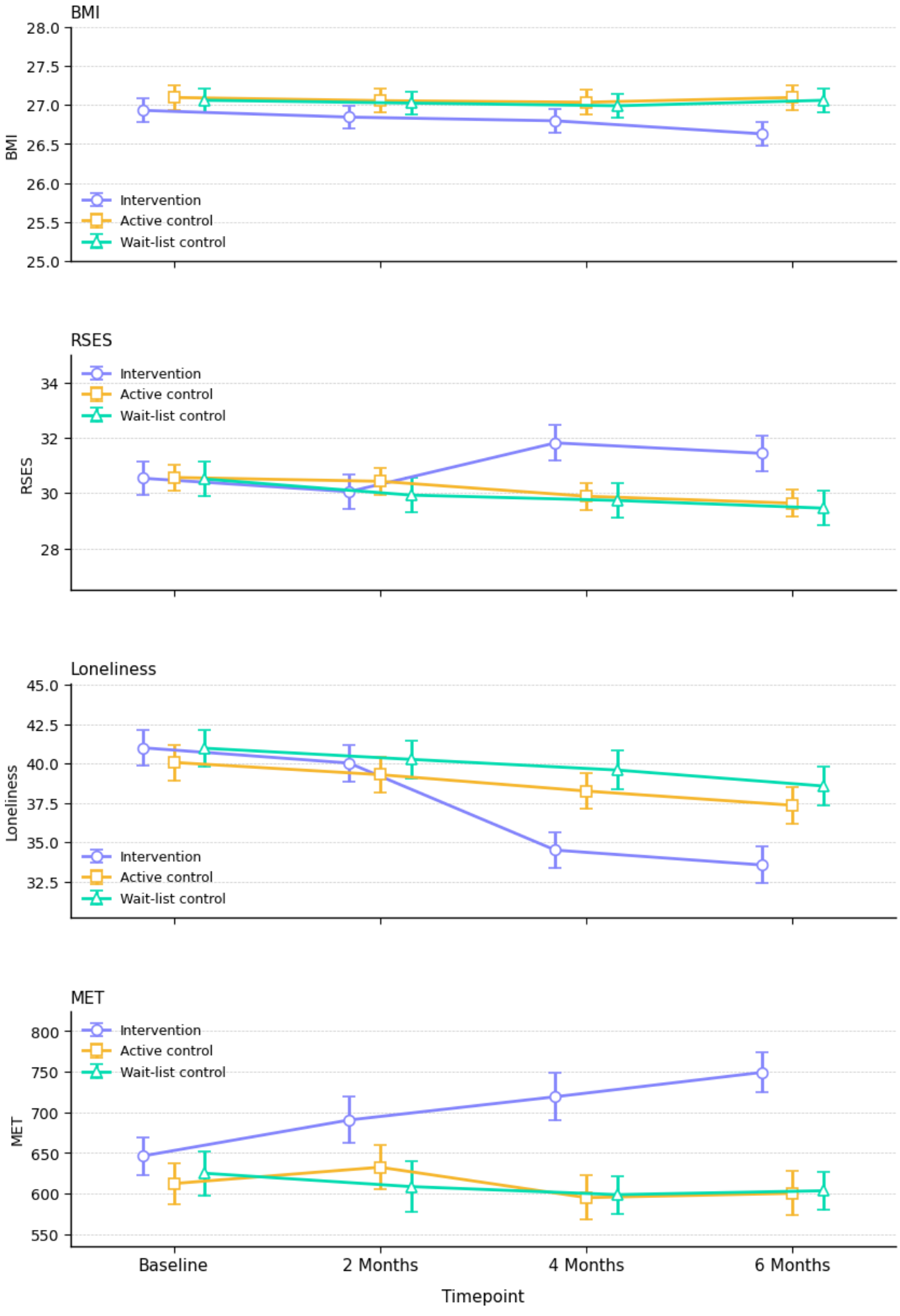
[l]Self‐Esteem: Rosenberg Self‐Esteem Scale.

[m]Loneliness: UCLA Loneliness Scale.

At 6 months, these benefits not only persisted but, in the case of BMI, became statistically significant. Although BMI differences were nonsignificant at 4 months, by 6 months the intervention group's mean BMI was 0.47 kg/m² lower than that of the active control (95% CI –0.89 to –0.04; *P*=.04) and 0.54 kg/m² lower than that of the waitlist control (95% CI –0.85 to –0.01; *P*=.03). Weekly MET‐minutes remained elevated among intervention participants (129.45 MET-minutes greater than active control; 95% CI 53.28-205.61, *P*<.001; 152.87 MET greater than waitlist control; 95% CI 68.35-237.38, *P*<.001). Similarly, self‐esteem differences persisted: intervention scores were 1.81 points higher than active control (95% CI 0.22-3.39; *P*=.02) and 1.99 points higher than waitlist control (95% CI 0.21-3.76; *P*=.03). Loneliness remained lower in the intervention arm, with mean scores 3.79 points below active control (95% CI –7.03 to –0.56; *P*=.03) and 5.02 points below waitlist control (95% CI –8.38 to –1.66; *P*<.001; Figure 2; Table 2).

Besides these intervention effects, no differences in mental health, specifically depression, anxiety, and stress, were observed between conditions, and no adverse or unintended events occurred in any study arm. Participants' evaluations of intervention quality did not differ between the intervention and control conditions, and follow‐up evaluations indicated high overall satisfaction rates among study participants.

Figure 2. Line charts of outcomes exhibiting significant changes over time in the intervention, active control, and waitlist control groups (mean, SE). MET: metabolic equivalent of task; RSES: Rosenberg Self-Esteem Scale.



Discussion

Principal Findings

In this RCT, a digitally delivered peer-support intervention yielded significant improvements in both physical and psychological outcomes among university students with preobesity. Compared with participants in the active control and waitlist control groups, those receiving digital peer support experienced a greater reduction in BMI over 6 months, accompanied by marked increases in weekly physical activity. Notably, the intervention also promoted significant enhancements in self-esteem and produced a meaningful decrease in loneliness scores. These convergent improvements demonstrate that a well-structured digital peer-support program can effectively address weight-related metrics while simultaneously bolstering mental well-being in a high-risk young adult population.

Comparison with Prior Work

When contextualized within the broader literature, our results both corroborate and extend existing findings. Prior meta-analyses of peer-support interventions have reported modest reductions in weight and BMI, often limited to decreases on the order of 0.5 - 1.0 kg and 0.1 - 0.2 kg/m², respectively [48]. In contrast, the BMI improvement observed here exceeded those averages, suggesting that delivering peer support through a comprehensive digital platform may enhance engagement and efficacy. Similarly, earlier work by Yeo and colleagues found that digital peer-support programs led to higher levels of physical activity [24]. Our trial confirms that young adults can achieve and maintain substantial activity gains when supported by peers in a virtual setting. More striking is the contrast between our findings on psychosocial outcomes and those of many prior weight management trials. Whereas some research has found limited or no improvement in self-esteem or loneliness following peer support, often attributing these null effects to insufficient social cohesion or lack of personalized engagement [8], our intervention clearly produced robust psychological benefits. The structure of our digital program, which combined interactive workshops, ongoing peer dialogues, gamified challenges, and personalized goal-setting, likely fostered a sense of community and accountability that translated into higher self-esteem and reduced perceived social isolation.

Potential Mechanisms

Several features of our study design and implementation may underlie these positive outcomes. First, we targeted college students with “preobesity,” a population at a critical inflection point that is often overlooked by both obesity-prevention initiatives and mental health programs [2]. By focusing on individuals with BMI between 25 and 29.9 kg/m², our intervention capitalized on a phase when behavior change is particularly feasible, as these students are frequently less entrenched in unhealthy habits and experience fewer physiological impediments compared to individuals already in the obese range [52]. Second, the digital nature of our peer-support platform enabled consistent, real-time interaction and social reinforcement, meeting students in their preferred communication channels. Daily micro-tasks, biweekly group challenges, and digital badges created an environment in which healthy behaviors became normalized and visible, thereby leveraging social comparison and accountability to strengthen motivation. Third, the inclusion of both an attention-matched active control and a waitlist control increased the rigor of our design. By demonstrating that the intervention’s benefits exceeded those of an equally engaging, wellness-focused program, we confirmed that the peer-support component itself was responsible for the superior outcomes, rather than the generic effects of participation or attention.

Behavioral and psychosocial theories provide mechanistic explanations for the observed effects. According to social cognitive theory, observing peers’ successes and receiving encouragement enhances self-efficacy, which in turn promotes adoption and maintenance of healthy behaviors [53]. In our study, participants saw fellow group members achieving step-count goals, sharing healthy meal ideas, and offering positive feedback, all of which contributed to a collective sense of possibility. As self-efficacy increased, participants likely felt more capable of meeting weekly activity targets and monitoring their weight, resulting in the higher MET scores and lower BMI values observed at 4 and 6 months. Concurrently, the online peer-support format fostered social connectedness and reduced feelings of isolation, with participants reporting a growing sense of belonging as they exchanged challenges and celebrated accomplishments in a respectful, nonjudgmental space [54]. Improvements in self-esteem likely stemmed from personal achievements, such as meeting fitness goals and from external validation like peer praise, thereby reinforcing a more positive self-concept [55]. Moreover, reductions in loneliness may have mitigated stress and negative affect, both of which are known barriers to sustained lifestyle change [54]. By alleviating emotional distress and enhancing social support simultaneously, the intervention created a synergistic environment in which mental well-being and physical health could rise together [56].

Implications

Finally, the integration of physical, behavioral, and psychosocial measures in our outcome assessment underscores the value of holistic intervention approaches. In contrast to many programs that focus narrowly on either weight loss or mental health, our trial demonstrates that digital peer support can generate parallel benefits across multiple domains. This alignment is particularly important in the university setting, where academic pressures, social transitions, and evolving peer networks intersect to shape students’ health behaviors and psychological states. By addressing these factors in concert, providing evidence-based guidance on nutrition and exercise alongside structured opportunities for emotional support, the intervention enabled participants to tackle weight management without sacrificing mental health, and vice versa. Such a comprehensive strategy not only enhances immediate outcomes, such as lower BMI, higher activity levels, improved self-esteem, and reduced loneliness, but also establishes a foundation for sustained long-term well-being. Given the extensive reach and accessibility of digital platforms, our findings underscore the importance of integrating peer-supported programming as a foundational element of university health promotion strategies.

Limitations

This study has several limitations. First, the use of self-reported weight and physical activity data may have introduced measurement error, given the well-established divergence between self-reported and objectively measured values. Second, relying on BMI as the sole physical outcome may not adequately capture changes in body composition, as BMI does not differentiate between fat and lean mass. Third, the study sample consisted exclusively of digitally literate, university students with preobesity, which may limit the generalizability of the findings to other populations, including individuals in different age groups, weight categories, or with limited digital access or engagement. Fourth, we observed no significant changes in some secondary outcomes. This null finding may be attributable to the relatively short follow-up period and limited sensitivity of the self-report instruments. Long-term follow-up is needed to determine whether improvements in BMI, physical activity, and mental health are sustained beyond the intervention and to evaluate any delayed effects on the incidence of obesity and mental health disorders. Fifth, we could not directly verify completion of some self-guided micro tasks at the individual level; future trials should incorporate objective sensing or device-based validation to confirm task completion without increasing participant burden.

Conclusions

In this RCT among university students with preobesity, a digital peer-support intervention led to improvements in both physical and mental health outcomes compared to a control condition. Participants receiving the program attained a modest reduction in BMI and increased physical activity, while also reporting higher self-esteem and lower loneliness than controls. These findings highlight the promise of interventions that leverage peer support via digital platforms, an accessible and replicable approach to addressing emerging health risks in young adults. The intervention's digital, peer-driven design enables broad dissemination across campus settings and beyond, engaging at-risk students who may not use traditional health services. Notably, this trial is among the first to target the preobesity stage with a holistic approach that concurrently addresses weight management and mental health. By intervening early with an integrated program, this approach prevents progression to obesity while mitigating psychological distress, illustrating a novel paradigm for young adult health promotion.

Acknowledgments

The authors thank all the participants for their time and involvement in this study, research assistants for their hard work throughout the project, and all the faculty members who helped coordinate and support the study. The authors further confirm that no generative artificial intelligence tools were used in the drafting or editing of this manuscript.

Funding

This work was supported by the Guangdong Provincial Sports Bureau Science & Technology Innovation and Sports Culture Development Program (GDSS2024N069) and by the Guangdong Provincial Education Science Planning Leading Group Office (Higher Education Special) (2024GXJK439).

Data Availability

The datasets generated or analyzed during this study are available in the Harvard Dataverse [57].

Authors' Contributions

Conceptualization: XL, TC, and RM
Methodology: RM and TL
Formal analysis: TM and TL
Data curation: TM
Investigation: XL and RM
Software and visualization: RM
Writing – original draft preparation: XL and RM
Writing – review and editing: XL and RM
Supervision: XL and TC

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File, 1362 KB - mhealth_v14i1e78960_app1.pdf]

References

1. Cole TJ, Lobstein T. Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. Pediatr Obes 2012 Aug;7(4):284-294. [doi: 10.1111/j.2047-6310.2012.00064.x] [Medline: 22715120]
2. Bailey CP, Sharma S, Economos CD, Hennessy E, Simon C, Hatfield DP. College campuses' influence on student weight and related behaviours: a review of observational and intervention research. Obes Sci Pract 2020 Dec;6(6):694-707. [doi: 10.1002/osp4.445] [Medline: 33354347]
3. Odlaug BL, Lust K, Wimmelmann CL, et al. Prevalence and correlates of being overweight or obese in college. Psychiatry Res 2015 May 30;227(1):58-64. [doi: 10.1016/j.psychres.2015.01.029] [Medline: 25770354]
4. Ilić M, Pang H, Vlaški T, Grujičić M, Novaković B. Prevalence and associated factors of overweight and obesity among medical students from the Western Balkans (South-East Europe Region). BMC Public Health 2024 Jan 2;24(1):29. [doi: 10.1186/s12889-023-17389-7] [Medline: 38166959]
5. Dipietro L, Zhang Y, Mavredes M, et al. Physical activity and cardiometabolic risk factor clustering in young adults with obesity. Med Sci Sports Exerc 2020 May;52(5):1050-1056. [doi: 10.1249/MSS.0000000000002214] [Medline: 31764468]
6. Emon HH, Sarker S, Lima MSA, et al. Prevalence of overweight and obesity and their impact on academic performance and psychological well-being among university students in 2024 in Bangladesh. In: Ahmed SK, editor. PLoS ONE 2024;19(12):e0315321. [doi: 10.1371/journal.pone.0315321] [Medline: 39656712]
7. Dakanalis A, Voulgaridou G, Alexatou O, et al. Overweight and obesity is associated with higher risk of perceived stress and poor sleep quality in young adults. Medicina (Kaunas) 2024 Jun 14;60(6):983. [doi: 10.3390/medicina60060983] [Medline: 38929600]
8. Streram S, Burrows T, Duncan MJ, Hutchesson M. Health behaviour interventions to improve mental health outcomes for students in the university setting: a systematic review of randomised controlled trials. Int J Behav Nutr Phys Act 2025 Mar 11;22(1):32. [doi: 10.1186/s12966-025-01718-7] [Medline: 40069770]
9. Lanoye A, Brown KL, LaRose JG. The transition into young adulthood: a critical period for weight control. Curr Diab Rep 2017 Oct 2;17(11):114. [doi: 10.1007/s11892-017-0938-4] [Medline: 28971312]
10. Vadeboncoeur C, Townsend N, Foster C. A meta-analysis of weight gain in first year university students: is freshman 15 a myth? BMC Obes 2015;2(1):22. [doi: 10.1186/s40608-015-0051-7] [Medline: 26217537]
11. West DS, Monroe CM, Turner-McGrievy G, et al. A technology-mediated behavioral weight gain prevention intervention for college students: controlled, quasi-experimental study. J Med Internet Res 2016 Jun 13;18(6):e133. [doi: 10.2196/jmir.5474] [Medline: 27296086]
12. Harrison S, Dixon P, Jones HE, Davies AR, Howe LD, Davies NM. Long-term cost-effectiveness of interventions for obesity: a Mendelian randomisation study. In: Veerman JL, editor. PLoS Med 2021 Aug;18(8):e1003725. [doi: 10.1371/journal.pmed.1003725] [Medline: 34449774]
13. Thorpe KE, Joski PJ. Estimated reduction in health care spending associated with weight loss in adults. JAMA Netw Open 2024 Dec 2;7(12):e2449200. [doi: 10.1001/jamanetworkopen.2024.49200] [Medline: 39636635]
14. Bjerregaard LG, Jensen BW, Ängquist L, Osler M, Sørensen TIA, Baker JL. Change in overweight from childhood to early adulthood and risk of type 2 diabetes. N Engl J Med 2018 Apr 5;378(14):1302-1312. [doi: 10.1056/NEJMoa1713231] [Medline: 29617589]
15. Gortmaker SL, Wang YC, Long MW, et al. Three interventions that reduce childhood obesity are projected to save more than they cost to implement. Health Aff (Millwood) 2015 Nov;34(11):1932-1939. [doi: 10.1377/hlthaff.2015.0631] [Medline: 26526252]
16. The Heavy Burden of Obesity: The Economics of Prevention: OECD; 2019. [doi: 10.1787/67450d67-en]
17. Dhar D, Packer J, Michalopoulou S, et al. Assessing the evidence for health benefits of low-level weight loss: a systematic review. Int J Obes (Lond) 2025 Feb;49(2):254-268. [doi: 10.1038/s41366-024-01664-7] [Medline: 39487296]
18. Pfisterer J, Rausch C, Wohlfarth D, Bachert P, Jekauc D, Wunsch K. Effectiveness of physical-activity-based interventions targeting overweight and obesity among university students-a systematic review. Int J Environ Res Public Health 2022 Aug 1;19(15):9427. [doi: 10.3390/ijerph19159427] [Medline: 35954789]
19. Cao B, Xu J, Li R, Teopiz KM, McIntyre RS, Chen H. Interventions targeting comorbid depression and overweight/obesity: a systematic review. J Affect Disord 2022 Oct 1;314:222-232. [doi: 10.1016/j.jad.2022.07.027] [Medline: 35878825]
20. Oftedal S, Fenton S, Hansen V, et al. Changes in physical activity, diet, sleep, and mental well-being when starting university: a qualitative exploration of Australian student experiences. J Am Coll Health 2024 Dec;72(9):3715-3724. [doi: 10.1080/07448481.2023.2194426] [Medline: 37014766]
21. Lee KM, Hunger JM, Tomiyama AJ. Weight stigma and health behaviors: evidence from the eating in America study. Int J Obes (Lond) 2021 Jul;45(7):1499-1509. [doi: 10.1038/s41366-021-00814-5] [Medline: 33934109]
22. Flint SW, Vázquez-Velázquez V, Le Brocq S, Brown A. The real-life experiences of people living with overweight and obesity: a psychosocial perspective. Diabetes Obes Metab 2025 Apr;27 Suppl 2(Suppl 2):35-47. [doi: 10.1111/dom.16255] [Medline: 39931901]
23. Jones RA, Lawlor ER, Birch JM, et al. The impact of adult behavioural weight management interventions on mental health: a systematic review and meta-analysis. Obes Rev 2021 Apr;22(4):e13150. [doi: 10.1111/obr.13150] [Medline: 33103340]

24. Yeo G, Fortuna KL, Lansford JE, Rudolph KD. The effects of digital peer support interventions on physical and mental health: a review and meta-analysis. Epidemiol Psychiatr Sci 2025 Feb 13;34:e9. [doi: 10.1017/S2045796024000854] [Medline: 39945388]
25. Ufholz K. Peer support groups for weight loss. Curr Cardiovasc Risk Rep 2020 Oct;14(10):19. [doi: 10.1007/s12170-020-00654-4]
26. Osborn TG, Town R, Ellis R, Buckman JEJ, Saunders R, Fonagy P. Implementing peer support in higher education: a feasibility study. SSM Ment Health 2022 Dec;2:100175. [doi: 10.1016/j.ssmmh.2022.100175] [Medline: 37916032]
27. D'Adamo L, Paraboschi L, Grammer AC, et al. Reach and uptake of digital mental health interventions based on cognitive-behavioral therapy for college students: a systematic review. J Behav Cogn Ther 2023 Jun;33(2):97-117. [doi: 10.1016/j.jbct.2023.05.002] [Medline: 37724304]
28. Toon J, Geneva M, Sharpe P, Lavin J, Bennett S, Avery A. Weight loss outcomes achieved by adults accessing an online programme offered as part of Public Health England's "Better Health" campaign. BMC Public Health 2022 Jul 30;22(1):1456. [doi: 10.1186/s12889-022-13847-w] [Medline: 35907834]
29. Stokols D. Establishing and maintaining healthy environments. Toward a social ecology of health promotion. Am Psychol 1992 Jan;47(1):6-22. [doi: 10.1037//0003-066x.47.1.6] [Medline: 1539925]
30. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
31. Craig CL, Marshall AL, Sjöström M, 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]
32. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
33. Zhang C, Wang T, Zeng P, et al. Reliability, validity, and measurement invariance of the General Anxiety Disorder Scale among Chinese Medical University students. Front Psychiatry 2021;12:648755. [doi: 10.3389/fpsyt.2021.648755] [Medline: 34093269]
34. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613. [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
35. Wang W, Bian Q, Zhao Y, et al. Reliability and validity of the Chinese version of the Patient Health Questionnaire (PHQ-9) in the general population. Gen Hosp Psychiatry 2014;36(5):539-544. [doi: 10.1016/j.genhosppsych.2014.05.021] [Medline: 25023953]
36. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5(1):63. [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
37. Fung SF. Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese university students. Health Qual Life Outcomes 2019 Mar 14;17(1):46. [doi: 10.1186/s12955-019-1113-1] [Medline: 30871563]
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
39. Lu W, Bian Q, Wang W, Wu X, Wang Z, Zhao M. Chinese version of the Perceived Stress Scale-10: a psychometric study in Chinese university students. In: Dang Y, editor. PLoS ONE 2017;12(12):e0189543. [doi: 10.1371/journal.pone.0189543] [Medline: 29252989]
40. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176. [doi: 10.1159/000376585] [Medline: 25831962]
41. Fung SF, Kong CYW, Liu YM, et al. Validity and psychometric evaluation of the Chinese version of the 5-item WHO Well-Being Index. Front Public Health 2022;10:872436. [doi: 10.3389/fpubh.2022.872436] [Medline: 35433612]
42. Rosenberg M, Schooler C, Schoenbach C, Rosenberg F. Global self-esteem and specific self-esteem: different concepts, different outcomes. Am Sociol Rev 1995 Feb;60(1):141. [doi: 10.2307/2096350]
43. Jiang C, Zhu Y, Luo Y, et al. Validation of the Chinese version of the Rosenberg Self-Esteem Scale: evidence from a three-wave longitudinal study. BMC Psychol 2023 Oct 18;11(1):345. [doi: 10.1186/s40359-023-01293-1] [Medline: 37853499]
44. Russell DW. UCLA Loneliness Scale (version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
45. Ip H, Suen YN, Hui LMC, Cheung C, Wong SMY, Chen EYH. Psychometric properties of the variants of the Chinese UCLA Loneliness Scales and their associations with mental health in adolescents. Sci Rep 2024 Oct 21;14(1):24663. [doi: 10.1038/s41598-024-75739-w] [Medline: 39433867]
46. Lee RM, Robbins SB. Measuring belongingness: the social connectedness and the social assurance scales. J Couns Psychol 1995;42(2):232-241. [doi: 10.1037//0022-0167.42.2.232]
47. Yoon E, Jung KR, Lee RM, Felix-Mora M. Validation of social connectedness in mainstream society and the ethnic community scales. Cultur Divers Ethnic Minor Psychol 2012 Jan;18(1):64-73. [doi: 10.1037/a0026600] [Medline: 22250899]
48. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688. [doi: 10.2196/19688] [Medline: 33988126]

49. Faul F, Erdfelder E, Lang AG, 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]
50. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
51. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1). [doi: 10.18637/jss.v067.i01]
52. Burgess E, Hassmén P, Pumpa KL. Determinants of adherence to lifestyle intervention in adults with obesity: a systematic review. Clin Obes 2017 Jun;7(3):123-135. [doi: 10.1111/cob.12183] [Medline: 28296261]
53. Islam KF, Awal A, Mazumder H, et al. Social cognitive theory-based health promotion in primary care practice: a scoping review. Heliyon 2023 Apr;9(4):e14889. [doi: 10.1016/j.heliyon.2023.e14889] [Medline: 37025832]
54. Bravata DM, Kim J, Russell DW, Goldman R, Pace E. Digitally enabled peer support intervention to address loneliness and mental health: prospective cohort analysis. JMIR Form Res 2023 Nov 6;7:e48864. [doi: 10.2196/48864] [Medline: 37930770]
55. Asadabadi FS, Karami K. The effect of exercise training videos on self-esteem of 7-10-year-old obese girls: a randomized controlled trial. BMC Public Health 2025 May 7;25(1):1684. [doi: 10.1186/s12889-025-22832-y] [Medline: 40336009]
56. Holt-Lunstad J. Social connection as a critical factor for mental and physical health: evidence, trends, challenges, and future implications. World Psychiatry 2024 Oct;23(3):312-332. [doi: 10.1002/wps.21224] [Medline: 39279411]
57. Efficacy of a digital peer support program on weight management and mental health in pre-obese university students: a randomized controlled trial. HARVARD Dataverse. URL: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/BZGDBB [accessed 2026-01-07]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
GAD-7: Generalized Anxiety Disorder Scale 7
MET: metabolic equivalent of task
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale
RCT: randomized controlled trial
SCS: Social Connectedness Scale
UCLA: University of California, Los Angeles Loneliness Scale
WEMWBS: Warwick-Edinburgh Mental Well-being Scale
WHO-5: 5-item World Health Organization Well-Being Index

Edited by L Buis; submitted 12.Jun.2025; peer-reviewed by D Tang, J Liu; revised version received 03.Oct.2025; accepted 16.Dec.2025; published 21.Jan.2026.

Please cite as:
Liu X, Liu T, Chen T, Ma R
Efficacy of a Digital Peer Support Program on Weight Management and Mental Health in University Students With Preobesity: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e78960
URL: https://mhealth.jmir.org/2026/1/e78960
doi:10.2196/78960

Original Paper

Analysis of Training Behavior in Users of a Fitness App: Cross-Sectional Study

Andrea Fuente-Vidal[1,2], PhD; Roger Prat[1], MSc; Juan Manuel Arribas-Marin[3,4], PhD; Oscar Bastidas-Jossa[5], PhD; Myriam Guerra-Balic[1], MD, PhD; Begonya Garcia-Zapirain[5], PhD; Joel Montane[1,6*], PhD; Javier Jerez-Roig[2,7,8*], PhD

[1]Faculty of Psychology, Education and Sports Sciences (FPCEE) Blanquerna, Universitat Ramon Llull, Barcelona, Spain

[2]Research Group on Methodology, Methods, Models and Outcomes of Health and Social Sciences (M3O), Faculty of Health Sciences and Welfare, Centre for Health and Social Care Research (CESS), Universitat de Vic - Universitat Central de Catalunya, Vic, Spain

[3]San Juan de Dios University School of Nursing and Physical Therapy, Comillas Pontifical University, Madrid, Spain

[4]San Juan de Dios Foundation, Madrid, Spain

[5]eVIDA Research Group, Universidad de Deusto, Bilbao, Spain

[6]School of Health Sciences, Blanquerna, Universitat Ramon Llull, Barcelona, Spain

[7]Institute of Sport Science and Innovations, Lithuanian Sports University, Kaunas, Lithuania

[8]Institute for Research and Innovation in Life and Health Sciences in Central Catalonia (IRIS-CC), Universitat de Vic - Universitat Central de Catalunya, Vic, Spain

[*]these authors contributed equally

Corresponding Author:
Joel Montane, PhD
Faculty of Psychology, Education and Sports Sciences (FPCEE) Blanquerna
Universitat Ramon Llull
Carrer del Císter, 34
Barcelona, 08022
Spain
Phone: 34 932 53 30 00
Email: joelmm@blanquerna.url.edu

Abstract

Background: Mobile health (mHealth) apps are increasingly being used to promote physical activity (PA) and can support exercise uptake and maintenance. Despite their potential, these tools face high dropout rates and inconsistent adherence, posing a significant challenge. Understanding how users engage with fitness apps is essential for improving user experience and health outcomes.

Objective: This study aims to analyze user behavior patterns in the Mammoth Hunters (MH) fitness app (Mammoth Hunters SL), focusing on retention (days from registration to user's last recorded training session), average weekly training frequency, and adherence (alignment between planned and actual training). We examined how these outcomes are influenced by sociodemographic, motivational, and other variables.

Methods: This cross-sectional study involved 2771 Mammoth Hunters app users. In a subsample (n=289), training data were complemented by motivational data acquired through online surveying via an ad-hoc scale (internal consistency >0.83) based on the self-determination theory (SDT). Descriptive statistics and nonparametric tests (Kruskal-Wallis, Dunn post-hoc, and Spearman correlation) were used to assess correlation between sociodemographic, motivation, and training behavior variables.

Results: Mean retention (days) was significantly higher among males than females (135 vs 109, respectively; $P<.01$), users in the subscription vs free plan (154 vs 81; $P<.001$), active or very active individuals vs inactive, midbuilt vs thin body types (132 vs 120; $P=.001$), and those with slightly lower BMI. Users pursuing antiaging or muscle gain goals showed longer retention than those aiming to lose weight (gain: 132, antiaging: 128, lose weight: 116; $P<.001$). Average weekly frequency (sessions per week) of training was statistically significantly different by sex (male: 1.9 vs female: 1.8; $P=.04$), body type (thin: 1.96 vs mid: 1.77; $P=.04$), activity level (very active: 2.05 vs inactive: 1.83; $P=.04$), and motivation type (extrinsic introjected motivation correlated positively: $r=0.17$; $P<.05$), but did not correlate with perceived difficulty or fitness goals. Adherence, defined as actual vs targeted training frequency, was only significantly different among body types, with thin users showing higher adherence than the midbuilt

group (57% vs 52.1%; *P*=.02). Intrinsic motivation showed a positive correlation with retention (*r*=0.19; *P*=.002), as did identified motivation (*r*=0.12; *P*<.05).

Conclusions: This study shows that retention is influenced by demographic factors, with males, subscribers, previously active, midbuilds, those aiming to gain muscle, and individuals with autonomous types (ie, intrinsic and identified) of motivation displaying greater long-term participation. These findings provide valuable preliminary insight into the complexities of exercise training behavior in apps. They suggest that training frequency, retention, and adherence do not respond to the same factors. App developers, researchers, and trainers should assess these variables separately and develop strategies accordingly.

(JMIR Mhealth Uhealth 2026;14:e72201) doi:10.2196/72201

KEYWORDS

fitness app; physical activity; exercise adherence; retention; motivation; mHealth

Introduction

Adherence to Physical Activity and Exercise

Physical activity (PA) and exercise are fundamental components of a healthy lifestyle, with well-established benefits for physical and mental well-being. PA, as defined by the World Health Organization (WHO), encompasses any bodily movement that results in energy expenditure, while exercise is considered a structured subset of PA, performed with the intent of improving or maintaining physical fitness [1]. Regular engagement in PA is crucial for reducing the risk of chronic diseases, yet adherence to recommended activity levels remains a global challenge [2]. Despite the widespread awareness of PA benefits, sustaining an active lifestyle is often hindered by behavioral, environmental, and psychological barriers [3]. Understanding factors that influence adherence is therefore critical for improving PA participation and ensuring long-term engagement.

Influence of mobile health on PA Behaviors

In recent years, mobile health (mHealth) apps have emerged as a potential solution to bridge the gap between PA recommendations and actual adherence. Fitness apps, a subset of mHealth, offer structured training programs, progress tracking, and personalized feedback, aiming to enhance motivation and user engagement. The widespread availability of smartphones has contributed to a surge in fitness app usage, with millions of users accessing digital exercise programs globally [4]. These apps incorporate behavior change techniques such as goal setting, social support, and gamification to facilitate sustained exercise habits [5]. However, despite their potential, high attrition rates and inconsistent long-term adherence pose significant challenges to their effectiveness [6]. Recent evidence reinforces that these barriers still persist across different populations and intervention designs. For example, several authors reported significant dropout rates even in gamified or socially incentivized fitness apps [7,8]. Similarly, previous studies highlighted continued adherence challenges in young individuals or older adults despite tailored mHealth interventions [6,7,9].

Adherence to exercise, particularly in digital interventions, remains a complex issue, often inconsistently defined across studies. Traditional adherence models typically assess exercise frequency, duration, and intensity, yet these criteria may not fully capture engagement in app-based fitness programs [3]. Furthermore, users may abandon apps due to technical difficulties, loss of motivation, or unrealistic expectations [10]. Consequently, understanding the determinants of fitness app adherence requires a multidimensional approach, integrating psychological, technological, and behavioral perspectives [11].

The challenges surrounding fitness app adherence are compounded by factors such as user characteristics, app usability, and the broader social and environmental contexts in which users engage with digital interventions. Studies have highlighted that individual attributes such as age, sex, health consciousness, and baseline PA levels may influence the likelihood of sustained engagement with fitness apps [12]. Additionally, app design elements, including intuitive navigation, feedback mechanisms, and interactive features, play a crucial role in user retention [13]. Social and motivational factors, such as competition, social support, and reinforcement strategies, have also been shown to impact adherence levels in digital exercise interventions [5]. Recent evidence highlights the importance of incorporating behavioral theories and enhancing usability and perceived value in reducing attrition and promoting sustained engagement with mHealth tools. For instance, recent findings emphasize the relevance of behavioral theories in crafting more effective mHealth interventions, showing how tailored features can reduce dropout and improve user retention [14]. Similarly, perceived value and usability have been identified as key drivers of long-term engagement with digital health tools [15]. Other studies suggest that personalization, motivational strategies, and social features are critical to increasing user commitment [12,16], overall highlighting the multifactorial nature of adherence and reinforcing the need for user-centered app design approaches.

Given the rising reliance on digital solutions for health and fitness, it is imperative to explore how different aspects of fitness apps contribute to sustained PA engagement. Previous research presents mixed findings on the long-term efficacy of fitness apps in promoting adherence, with some studies reporting positive behavior changes and others limited long-term impact [8,17].

Understanding real users' training behavior, beyond theoretical frameworks or self-reported intentions, is essential to identify how engagement translates into real-world usage. Analyzing app usage data provides evidence of behavior patterns, allowing researchers to identify which user profiles are more likely to sustain app use. This study establishes a theoretical and practical differentiation between adherence and retention, and how they

relate to user motivation to exercise. This is a critical matter, since sustained usage is key to ensuring the long-term impact of digital health interventions [18].

Study Goal

This study aims to explore the factors influencing the training behavior of users of a fitness app, focusing specifically on exercise adherence, retention, and motivation, and to explore how these outcomes are influenced by sociodemographic, motivational, and training-related factors. We hypothesize that sociodemographic characteristics, motivation types, and training behaviors significantly influence users' retention, adherence, and frequency of training. Understanding these factors is essential for optimizing fitness app design, improving intervention strategies, and ultimately promoting long-term participation in exercise.

Methods

Data Collection and Processing

Data were collected in collaboration with Mammoth Hunters (MH; Mammoth Hunters SL), a fitness app that focused on high-intensity interval exercises to improve strength, endurance, and mobility. MH delivered structured programs rooted in functional movement, making it an ideal platform for investigating digital fitness adherence, motivation, and retention. MH was launched in 2014 by a team of fitness experts and scientists from Barcelona, Spain. A free version with limited access to certain workouts and features was available upon registration, while the Pro version (per subscription) provided full access to personalized plans, a greater variety of workouts, and advanced tracking tools. MH ceased operations in September 2021, being one of the most widely used high-intensity training apps worldwide, having accumulated a total of 719,421 users. The company's shutdown, as well as its noninvolvement in any of the stages of study, ensured no conflict of interest.

The study used a cross-sectional design. Data in the MH app database included user registries ranging from November 21, 2020, to May 27, 2022. Motivation data were collected via online surveying on March 20, 2022. Data cleaning and descriptive analyses were conducted using the R programming language (version 4.3.1; R Core Team) in the R Studio environment software (version 2023.9.1.494; Posit, PBC).

The initial MH dataset contained 5858 entries, which corresponded to users who had granted informed consent to share their deidentified data for analysis. To ensure the accuracy and relevance of the data, several cleaning steps were executed, including a convenience selection of the most relevant variables and exclusion of registries with insufficient or missing data (see Table S1 in Multimedia Appendix 1 for more detail of the MH app's original variables).

Outliers were identified through data visualization and consultation of descriptive statistics. A decision was made to remove all outliers to avoid distortion in the analysis, based on two main reasons: (1) certain tests run by MH staff members had intentionally introduced outlier scores to facilitate their identification and removal, and (2) some outliers resulted from the arbitrary temporal cutoff applied to the dataset, specifically, some participants had only just begun training with the app shortly before the data extraction date, leading to unrealistic extreme values in some outcome variables (eg, extremely low retention or extremely high adherence). Therefore, all values exceeding Q3+1.5×IQR or falling below Q1–1.5×IQR were removed.

The motivation-related outcome variables were determined using the means of the composite scores from the observable items corresponding to each factor in the scale, allowing us to define the latent variables. Outcome variables retention, frequency, and adherence were derived from existing variables on the mobile app (eg, number of sessions executed, last training date, and user sign-up date) to enhance analytical depth. User retention was calculated as the number of days between a user's initial registration date within the fitness app and their last recorded training session. Weekly training frequency was calculated by dividing the total number of training sessions completed by a user by the total number of weeks from their first to their last executed session. Adherence was quantified as the percentage to which a user's actual average weekly training frequency aligned with their initial plan (self-declared upon registration). Additionally, data types were adjusted as required to ensure compatibility and accuracy.

Following the described data cleaning steps, the final dataset comprised 2771 user entries. See Table 1 for more details on study variables.

Table 1. Description of explanatory and outcome variables.

Explanatory variables	Type	Description
Sex	Categorical; sociodemographic	User biological sex, with 2 categories: female and male.
Body type	Categorical; sociodemographic	Self-reported body type selected by the user at registration, out of 3 available categories: thin, mid, and strong.
Activity level	Categorical; sociodemographic	User activity level at the time of registration in the fitness app, with 3 categories: inactive, active, and very active.
Fitness goal	Categorical; sociodemographic	The goal the user aims to achieve through app use (selected from 3 available categories: lose weight, gain muscle, and antiaging).
Pro version	Categorical; training	Indicative of app user being subscribed to a payment ("Pro") program or not. Two categories: yes and no.
Training schedule	Categorical; training	Time of day in which the user executes most (>50%) of their training sessions. Processed into 3 categories: morning (5:30-12:30 hours), afternoon (12:31-20 hours), and night (20-5:29 hours).
Age	Numerical; sociodemographic	User's reported age at registration.
BMI	Numerical; sociodemographic	User BMI calculated from their declared height and weight.
Subjective body fat	Numerical; sociodemographic	Users' self-reported body fat.
Difficulty	Numerical; training	Average perceived exertion reported at the end of the training session. 0 (lowest)-10 (highest).
Enjoyment	Numerical; training	Average user-reported enjoyment after each training session. 0 (lowest)-5 (highest).
SDT[a]-based variables		
Intrinsic motivation	Numerical; motivation	Average score of the intrinsic motivation items on the scale: a decimal number between 1 (lowest) and 5 (highest).
Identified extrinsic motivation	Numerical; motivation	Average score of the identified extrinsic motivation items on the scale: a decimal number between 1 (lowest) and 5 (highest).
Introjected extrinsic motivation	Numerical; motivation	Average score of the introjected extrinsic motivation items on the scale: a decimal number between 1 (lowest) and 5 (highest).
Outcome variables		
Retention	Numerical; training	Measured as the total number of days from the user registration date in the app to their last recorded training session.
Frequency, weekly average	Numerical; training	Calculated by dividing the total number of user sessions by the total number of weeks between their first and last recorded sessions.
Adherence	Percentage; training	Defined as the percentage alignment between the user's actual weekly training frequency and their predefined weekly training goal.

[a]SDT: self-determination theory.

Motivational data, which had been previously collected (March 20, 2022) by means of an ad-hoc scale (Table S2 in Multimedia Appendix 1), provided insight as to the motivational regulation of a subsample (n=753) of MH users. The scale was based on the self-determination theory (SDT) [19,20]. It showed good fit indices and a 3-factor structure as confirmed per exploratory and confirmatory factor analyses, with internal consistency indices >0.830 for the 3 subscales (intrinsic, identified extrinsic, and introjected extrinsic motivations). Data obtained through surveying (n=753) and data obtained from the MH fitness app (n=2771) were then merged, and a sample consisting of n=328, for which both training and motivational data were available, was obtained. Thirty-nine registries had to be disregarded due to missing data for the calculation of adherence and weekly training frequency outcome variables. A resulting total of 289 was complete for all explanatory and outcome study variables.

Descriptive Analysis for Sociodemographic, Training, and Motivation Variables

Following data cleaning, descriptive statistics were computed to summarize and describe the dataset's characteristics. Frequencies and percentages were calculated for categorical variables to provide an overview of their distribution and proportions within the sample. For numerical variables, measures of central tendency (mean and median) and measures of dispersion (minimum, maximum, and quartiles) were obtained to characterize data distribution and its variability.

Inferential Analysis of Explanatory Variables

Normality was assessed using the D'Agostino-Pearson test, along with skewness and kurtosis coefficients to quantify distributional properties. Additionally, histograms and quantile-quantile plots were inspected to visually evaluate deviations from normality. The results indicated significant departures from normality, and nonparametric tests were used for subsequent analyses. The Kruskal-Wallis test was conducted to evaluate differences among groups, followed by post-hoc analysis using the Bonferroni correction to adjust for multiple comparisons. The effect size was assessed using Dunn, which quantifies the magnitude of observed differences. To examine relationships between numeric variables, Spearman correlation tests were performed. The Holm correction was applied to control for multiple comparisons and to adjust the significance levels accordingly. For training behavior analysis, the categorical variables evaluated were sex, Pro version, self-declared level of previous PA, body type, fitness goal, and training schedule. Additionally, explanatory numerical variables included age, BMI, subjective body fat, perceived difficulty, and enjoyment. Intrinsic, identified extrinsic, and introjected extrinsic motivations were also considered (Table 1). The relationship of all these variables was analyzed with three outcome variables: adherence, frequency, and retention.

Inferential Analysis of Outcome Variables

Adherence was calculated as the percentage match between the target weekly frequency, as selected by the user upon sign-up, and the actual, executed weekly training frequency. The latter was averaged by dividing the total number of executed sessions by the total number of weeks from the sign-up date to the last executed session for the given user. Finally, retention was measured as the total number of days from the user sign-up date to the user's last recorded training session (Table 1). Outcome variables were analyzed through the same procedures as explanatory variables.

Ethical Considerations

Ethical approval for this study was obtained from the Research Ethics Committee of Universitat Ramon Llull in March 2020 (reference code 1920003P). All included users provided informed consent to use their data for research purposes, either through the app at registration or through the motivational survey. All data were anonymized before analysis, ensuring the privacy and confidentiality of participants in compliance with data protection regulations. No compensation was provided to participants, as the data were collected retrospectively and only for research purposes.

Results

Descriptive Results for Sociodemographic, Training, and Motivation Variables

Our sample consisted of 2771 MH users. Of them, a 64.8% majority identified as male, and 35.2% as female. Their age range spanned from 21 to 64 years, with a median age of 43 years and a mean age of 42.45 years. Users' fitness goals varied, with the largest segment (46.6%) aiming to "gain muscle" mass, followed by those wanting to "lose weight" at 32%. A smaller portion, 21.4%, pursued "antiaging" benefits. Table 2 provides full details on sample description and other results.

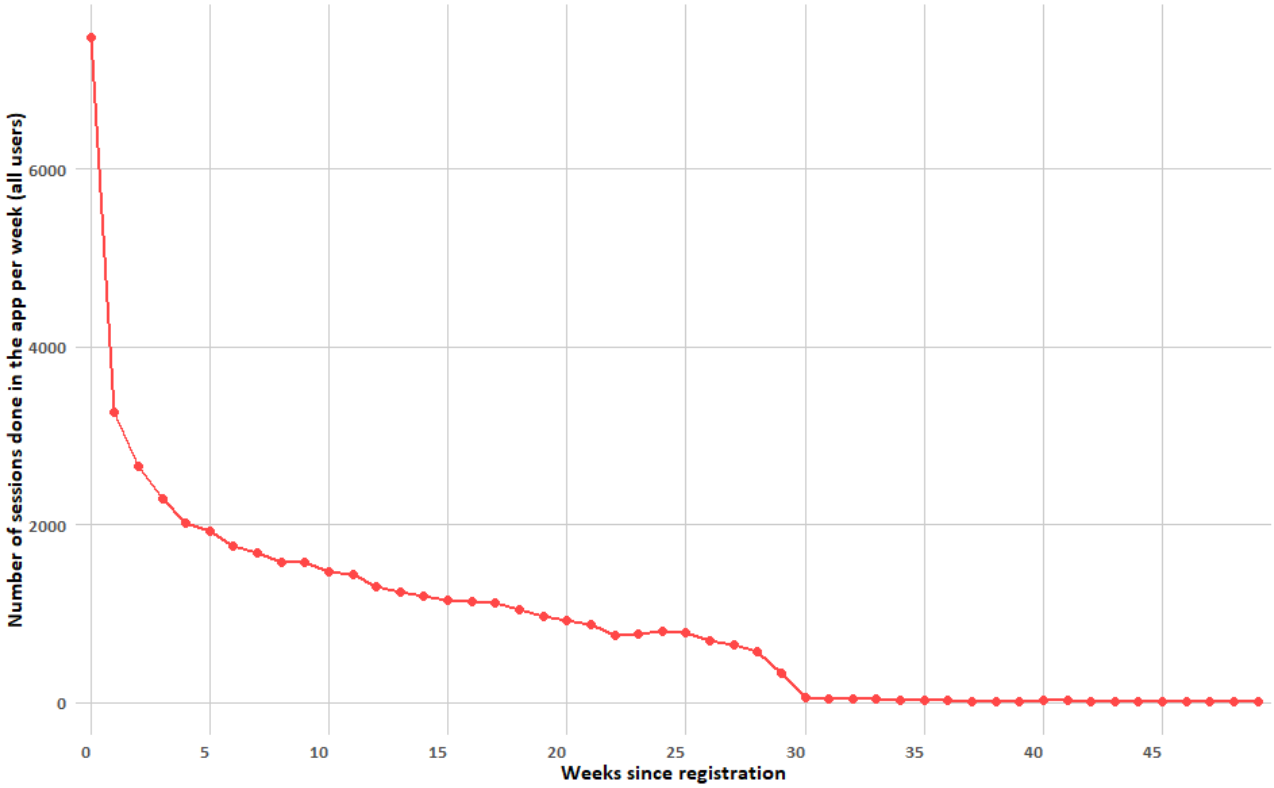
Table 2. Descriptive characteristics of the Mammoth Hunters user sample (N=2771).

Variable	Values
Sex, n (%)	
Male	1796 (64.8)
Female	975 (35.2)
Age (years)	
Mean (SD)	42.45 (9.8)
Median (IQR)	43 (21)
Range (minimum-maximum)	21-64
Body type, n (%)	
Thin	1269 (45.8)
Midbuild	1419 (51.2)
Strong	191 (6.9)
Body fat	
Mean (SD)	22.2 (6.4)
Median (IQR)	20 (14)
Range (minimum-maximum)	6-40
BMI	
Mean (SD)	23.46 (2.9)
Median (IQR)	23.44 (3.8)
Subscription type, n (%)	
Pro (paid) users	1726 (62.3)
Standard (free) users	768 (27.7)
Physical activity level, n (%)	
Active	1640 (59.2)
Very active	454 (16.4)
Inactive	675 (24.4)
Actual training schedule, n (%)	
Morning	928 (33.5)
Afternoon	1372 (49.5)
Night	471 (17)
Training difficulty (1-10)	
Mean (SD)	5.56 (1.78)
Median (IQR)	5.40 (2.10)
Range (minimum-maximum)	1-9.5
Enjoyment (1-5)	
Mean (SD)	3.58 (0.82)
Median (IQR)	3.50 (1.20)
Motivation (1-5)	
Intrinsic	
Mean (SD)	4.01 (0.74)
Median (IQR)	4 (0.90)
Range (minimum-maximum)	1.5-5
Identified extrinsic	

Variable	Values
Mean (SD)	4.42 (0.50)
Median (IQR)	4.67 (0.67)
Range (minimum-maximum)	3-5
Introjected extrinsic	
Mean (SD)	4.22 (0.61)
Median (IQR)	4.33 (1)
Range (minimum-maximum)	1-5
Retention (days)	
Mean (SD)	125.99 (92.60)
Median (IQR)	132.72 (144.36)
Range (minimum-maximum)	3.49-410.82
Training frequency (sessions per week)	
Mean (SD)	1.87 (1.52)
Median (IQR)	1.62 (1.84)
Range (minimum-maximum)	0.07-6.59
Adherence (%)	
Mean (SD)	54.24 (32.81)
Median (IQR)	47.31 (52.81)
Range (minimum-maximum)	1.18-166.67

Figure 1 illustrates the number of training sessions completed each week, by the total number of users, over a 49-week span. At the beginning, there was a sharp peak in the number of training sessions, with 7469 sessions recorded in the first week after user registration, for a total of 2771 users. Following this peak, the number of sessions decreased rapidly over the next several weeks. It declined to 2295 by the end of the first month (a reduction of 69.3%), to 1678 by the end of the second month, and to 1448 (a reduction of 80.6%) by the end of the third. By around week 10, the decline began to stabilize, though a gradual downward trend persisted. By the 30th week, the number of sessions plateaued at a much lower level, approximately below 100 sessions in a week.

Figure 1. Total number of sessions in time (per training week) for all users (N=2771).



The time from user registration to their first training session ranged from 0 to 319.84 days, with a median delay of 7.98 days and a mean of 30 days. Most users initiated training within the first few days after registering, and the frequency declined steeply after 10 days. Delays beyond 50 days were rare, and only a very small proportion of users waited more than 100 days. A small subgroup of 31 users showed exceptionally long delays between 200 and 350 days.

Differences Between Training Outcome Variables

Retention Results

All categorical variables except for training schedule showed statistically significant differences to retention (Table 3). For the sex variable, retention values were statistically significantly (P<.001) higher in the “male” group when compared to the “female” group, though the effect size was small (Dunn r=0.14). For the Pro version variable, indicative of whether the user was or was not subscribed for service at the time of data download, retention values were statistically significant (P<.001) with a moderate-to-large effect size (Dunn r=0.41), higher in the “yes” group than in the “no” group.” In regard to activity level, retention values were statistically significant (P<.001, Dunn r=0.11; and P<.001, Dunn r=0.07, respectively) and were higher for the “active” and “very active” groups than for the “inactive” group. No statistically significant (P>.05, Dunn r=0.03) differences were found in retention values between the “active” and “very active” groups. Refer to Figure 2 for further details on the Kruskal-Wallis results for the outcome variables.

Figure 2. Summary of Kruskal-Wallis findings for each outcome variable.

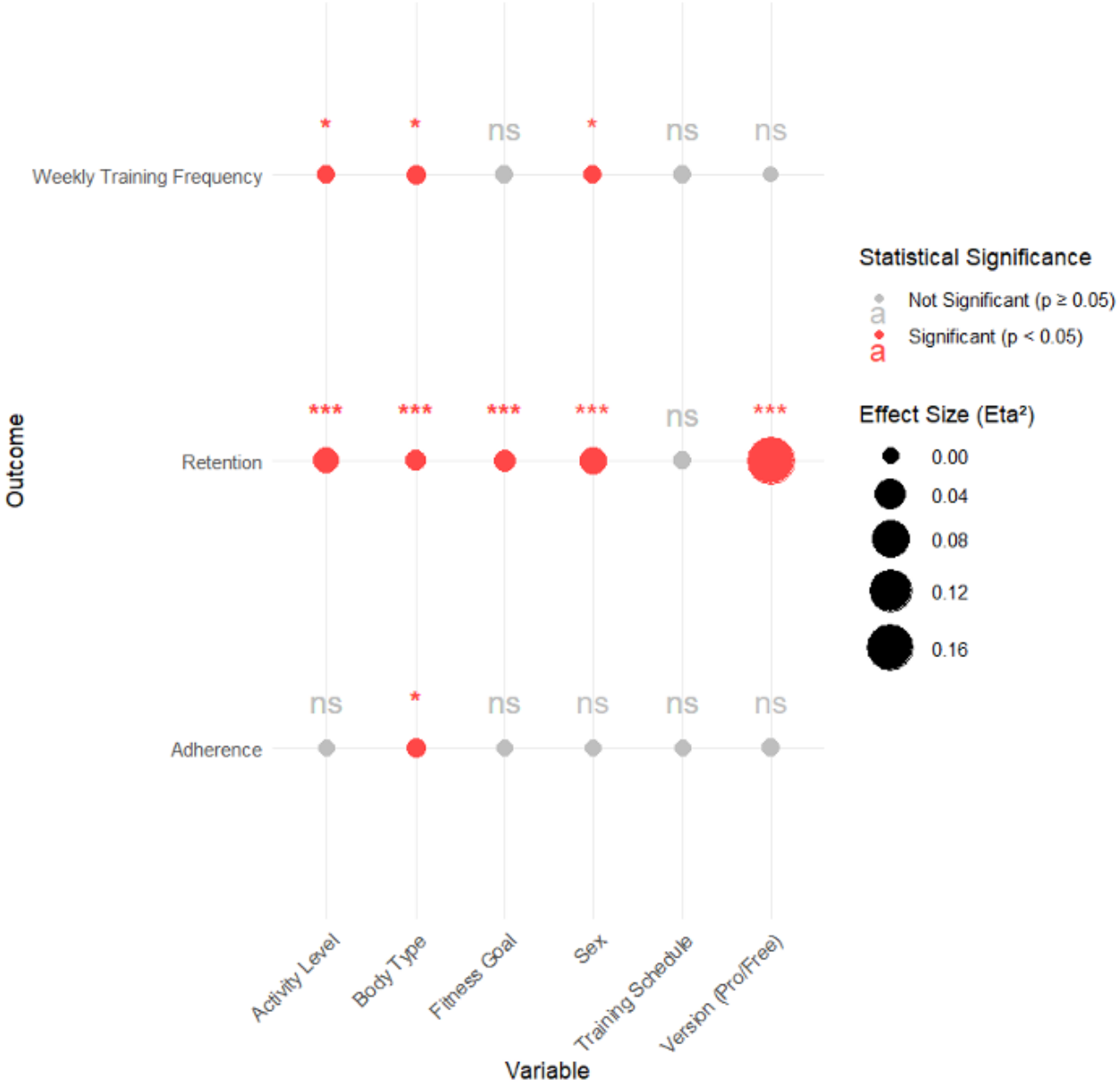


Table 3. Retention (in days), relative to categorical variables levels according to the multivariate analysis of variance (n = 2771).

Variable and group	Mean (SD)	H[a]	η^2 [b]	*P* [c]
Sex		56.151	0.019	<.001[d]
Female	109.43 (75.51)			
Male	134.99 (73.86)			
Pro version		459.965	0.163	<.001[d]
No	80.73 (70.11)			
Yes	153.53 (64.47)			
Activity level		47.414	0.016	<.001[d]
Inactive	109.74 (74.84)			
Active	130.69 (75.65)			
Very active	139.72 (69.57)			
Body type		14.303	0.004	.001[e]
Mid	132.37 (74.16)			
Thin	120.16 (76.27)			
Strong	117.44 (75.63)			
Fitness goal		21.982	0.007	<.001[d]
Antiaging	128.45 (70.23)			
Gain	132.15 (76.62)			
Lose	115.8 (75.18)			
Training schedule		3.613	0.000	.16
Afternoon	125.51 (74.23)			
Morning	128.49 (76.54)			
Night	119.07 (77.71)			

[a]H: Kruskal-Wallis H value.

[b]η2: eta squared.

[c]*P:* Kruskal-Wallis significance.

[d]*P*<.001.

[e]*P*<.01**.**

For the body type variable, the “mid” group retention values were statistically significantly higher (*P*<.002; Dunn *r* 0.07) than those in the “thin” group. No statistically significant differences were found in retention values between the “mid” and “strong” groups, nor between the “strong” and “thin” groups (*P*>.05; Dunn *r*=0.04 and *P*>.05; Dunn *r*=0.00, respectively). In the fitness goal variable, retention values were significantly lower for the “lose weight” group compared to the “antiaging” and “gain muscle” groups (*P*<.02; Dunn *r*=0.05 and *P*<.001; Dunn *r*=0.09, respectively) (Table 4).

Table 4. Post-hoc comparison of retention by categorical variable levels.

Variable	Comparison	U[a]	*P* value
Sex	• Female-male	• –7.493	• <.001b
Pro version	• No-yes	• –21.447	• <.001b
Activity level	• Active-inactive • Active-very active • Inactive-very active	• 6.067 • –1.796 • –5.641	• <.001b • .22 • <001b
Body type	• Mid-strong • Mid-thin • Strong-thin	• 2.034 • 3.579 • –0.234	• .13 • <.001c • >.99
Fitness goal	• Antiaging-gain • Antiaging-lose • Gain-lose	• –0.607 • 2.858 • 4.597	• >.99 • .01d • <.001b
Training schedule	• Afternoon-morning • Afternoon-night • Morning-night	• –1.062 • 1.266 • 1.850	• .87 • .62 • .19

[a]*U:* Standardized test statistic.

[b]***P*<.001**.

[c]*P*<.01.

[d]*P*<.05.

Regarding explanatory numerical variables, all of them showed statistically significant correlations with retention. Age had a moderate positive correlation (r=0.21; P<.001) with retention, while subjective body fat and BMI showed a low negative correlation (r=–0.13; P<.001; r=–0.06; P<.01, respectively). Training difficulty had a moderate, positive correlation (r=0.24; P<.001), and enjoyment had a low, positive correlation (r=0.11; P<.001).

Of all motivation dimensions, intrinsic motivation had the highest positive correlation (r=0.19; P<.01) with retention. Identified extrinsic motivation had a small, statistically significant, positive correlation (r=0.12; P<.05). Introjected extrinsic motivation had a much lower, positive, and nonstatistically significant correlation (r=0.07; P>.05; Table 5).

Table 5. Results of the correlation tests for motivation variables (N=289).

Variable	r [a]	*P* value
Retention		
INTRINS[b]	0.19	.01[c]
IDENT, extr.[d]	0.12	<.05[c]
INTROJ, extr.[e]	0.07	.22
Frequency, weekly		
INTRINS	0.001	>.99
IDENT, extr.	0.008	>.99
INTROJ, extr.	0.168	.01[c]
Adherence		
INTRINS	–0.002	>.99
IDENT, ext.	0.021	>.99
INTROJ, ext.	0.138	.06

[a]*r:* Spearman correlation coefficient.

[b]INTRINS: Intrinsic motivation.

[c]*P*<.05.

[d]IDENT, extr.: Identified extrinsic motivation.

[e]INTROJ, extr.: Introjected extrinsic motivation.

Average Weekly Frequency Results

Weekly frequency was found to be statistically significantly associated with sex, activity level, and body type (*P*<.05) and not significantly related to the Pro version, fitness goal, or training schedule (Table 6).

Table 6. Average weekly training frequency relative to categorical variables levels. Multivariate analysis of variance (N=2771).

Variable	Mean (SD)	H[a]	η^2 [b]	*P* [c]
Sex		3.950	0.001	.04[d]
Female	1.82 (1.35)			
Male	1.9 (1.29)			
Pro version		0.122	-0.001	.78
No	1.94 (1.45)			
Yes	1.83 (1.21)			
Activity level		6.199	0.001	.04[d]
Inactive	1.83 (1.34)			
Active	1.85 (1.29)			
Very active	2.05 (1.31)			
Body type		8.324	0.002	.02[d]
Mid	1.77 (1.23)			
Thin	1.96 (1.37)			
Strong	2.02 (1.39)			
Fitness goal		4.363	<0.001	.11
Antiaging	1.79 (1.28)			
Gain	1.91 (1.28)			
Lose	1.85 (1.36)			
Training schedule		4.519	0.001	.10
Afternoon	1.82 (1.30)			
Morning	1.93 (1.32)			
Night	1.94 (1.30)			

[a]H: Kruskal-Wallis H value.

[b]$\eta 2$: eta squared.

[c]*P*: Kruskal-Wallis significance.

[d]$P<.05$.

For the sex variable, figures in the "male" group were statistically significant ($P<.005$; Dunn $r=0.04$) and higher than those for the "female" group. In the activity level variable, the "very active" group values were significantly higher ($P<.05$; Dunn $r=0.05$) than those in the "inactive" group. Neither the "active" versus "inactive" groups nor the "active" versus "very active" groups showed any statistically significant differences in weekly training frequency values. For the body type variable, values in the "thin" group were statistically significantly ($P<.04$; Dunn $r=0.05$) higher than those of the "mid" group. No statistically significant differences were found between the "mid" and "strong" groups, nor between the "strong" and "thin" groups, for weekly training frequency values. For more detail on multivariate analyses of variance results for weekly training frequency, please refer to Table 7.

Table 7. Post-hoc comparison of average weekly training frequency by categorical variable levels.

Variable	Comparison	U [a]	P value
Sex	• Female-male	• –1.988	• .04b
Pro version	• No-yes	• 0.349	• .73
Activity level	• Active-inactive • Active-very active • Inactive-very active	• 0.599 • –2.218 • –2.427	• >.99 • .08 • .04b
Body type	• Mid-strong • Mid-thin • Strong-thin	• –1.989 • –2.501 • 0.727	• .14 • .04b • >.99
Fitness goal	• Antiaging-gain • Antiaging-lose • Gain-lose	• –1.716 • –0.372 • 1.679	• .26 • >.99 • .28
Training schedule	• Afternoon-morning • Afternoon-night • Morning-night	• –1.824 • –1.508 • –0.379	• .20 • .40 • >.99

[a]U = standardized test statistic.

[b]P<.05.

After corrections for multiple comparisons, none of the explanatory numerical variables reflected statistically significant correlations with weekly training frequency.

Regarding types of motivation, only introjected extrinsic motivation correlated significantly with training frequency *(*r=0.17; P<.05), after corrections for multiple comparisons, with a moderate-low positive correlation. Intrinsic and identified extrinsic motivations showed close to no correlation with frequency (r=0.00 and P>.05 for both) (Table 5).

Adherence Results

Adherence was found to be associated with one of the categorical variables (Table 8). These differences were found in body type, which presented a highly statistically significant (P<.005; Dunn r=–0.06) difference between the groups “mid” and “thin”. In particular, the values for the “mid” group were statistically significantly lower than those for the “thin” group. No statistically significant differences were found between groups “mid” and “strong,” nor between “strong” and “thin,” in adherence to the training program.

Table 8. Adherence, relative to categorical variables levels. Multivariate analysis of variance (N=2771).

Variable and group	Mean (SD)	H[a]	η^2 [b]	*P* [c]
Sex		1.479	0.000	.22
Female	54.02 (38.07)			
Male	54.35 (34.88)			
Pro version		3.260	0.000	.07
No	57.35 (39.85)			
Yes	52.35 (33.37)			
Activity level		2.394	0.000	.30
Inactive	53.76 (38.29)			
Active	53.91 (35.09)			
Very active	57.01 (35.39)			
Body type		7.322	0.002	.03[d]
Mid	52.1 (35.28)			
Thin	56.96 (37.32)			
Strong	52.01 (31.57)			
Fitness goal		2.960	0.000	.23
Antiaging	52.52 (34.7)			
Gain	55.34 (35.53)			
Lose	53.48 (37.37)			
Training schedule		2.843	0.000	.24
Afternoon	53.12 (35.66)			
Morning	56.05 (36.95)			
Night	53.61 (34.38)			

[a]*H*: Kruskal Wallis H value.

[b]$\eta 2$: eta squared.

[c]*P:* Kruskal-Wallis significance.

[d]$P<.05$.

No statistically significant differences were found between adherence and any of the groups in any of the remaining categorical variables (Table 9).

Table 9. Post-hoc comparison of adherence by categorical variables levels.

Variable	Comparison	U [a]	P value
Sex	• Female-male	• –1.216	• .22
Pro version	• No-yes	• 1.806	• .07
Activity Level	• Active-inactive • Active-very active • Inactive-very active	• 0.832 • –1.082 • –1.539	• >.99 • .84 • .37
Body Type	• Mid-strong • Mid-thin • Strong-thin	• –0.517 • –2.703 • –0.836	• >.99 • .02b • >.99
Fitness Goal	• Antiaging-gain • Antiaging-lose • Gain-lose	• –1.282 • –0.096 • 1.499	• .60 • >.99 • .40
Training schedule	• Afternoon - Morning • Afternoon - Night • Morning - Night	• –1.680 • –0.558 • 0.453	• .28 • >.99 • >.99

[a]U: Mann-Whitney standardized test statistic.

[b]P<.05.

For explanatory numerical variables, none reflected statistically significant correlations with adherence. Similarly, no types of motivation showed statistically significant correlations with adherence. Introjected extrinsic motivation rendered the highest (r= 0.14; P>.05) positive correlation, while intrinsic and identified extrinsic motivation showed little to no correlation (r=–0.00; P>.05, and r=0.02; P>.05, respectively; Table 5).

Summary of All Findings From the Inferential Analysis

A summary of all findings from the inferential analysis is presented in Table 10.

Table 10. Summary of statistically significant findings of the post-hoc inferential analyses.

Variable	Retention	Weekly training frequency	Adherence
Sex	Male>female	Male>female	—[a]
Pro version	Yes>no	—	—
Activity level	Active>inactive Very active>inactive	Very active>inactive	—
Fitness goal	Antiaging>lose Gain>lose	—	—
Body type	Mid>thin	Thin>mid	Thin>mid
Training schedule	—	—	—
Age	Positive correlation	—	—
BMI	Negative correlation	—	—
Subjective body fat	Negative correlation	—	—
Difficulty	Positive correlation	—	—
Enjoyment	Positive correlation	—	—
Intrinsic motivation	Positive correlation	—	—
Identified extrinsic motivation	Positive correlation	—	—
Introjected extrinsic motivation	—	—	—

[a]Not applicable.

Discussion

Summary of Key Findings

This study examined training behaviors among users of the MH fitness app and identified key factors associated with training behavior. Variables including adherence and retention were evaluated, with the latter having shown greater relevance in users long-term maintenance of training behavior. The main findings pointed at paid subscription and intrinsic motivation as being the most determinant factors to user retention. Other variables that correlated with retention included sex, body type, BMI, and fitness goal. In contrast, adherence was only linked to body type, while training frequency varied slightly by sex, activity level, motivation, and body type.

This piece of research involved 2771 individuals and is possibly one of the largest existing cohort studies of fitness app users to date. Previous large cohorts include the Konstanz Life Study with 1236 users of either fitness or nutrition apps [21]. Some systematic reviews have covered samples of 3555 participants from a total of 22 interventions (n=833 in the largest single study) [8] or 1622 total participants from 6 different studies [7]. Our work possibly also covers the longest time duration (18 months). Previous research has been 2-24 weeks [7], up to 5 months [22], or even 6 months in some cases [8].

Our sample figures fall within the "expected" ranges for a fitness app that offers high-intensity training, delivered electronically. Results are also in line with the systematic review by Stecher et al [8], which included participants between 10.6 and 61.5 years of age and found a mean of 39.6 (SD 6.5) years. Participants in this study presented some features worth noting, which were probably specific to our sample population. The majority (75.6%) of them were previously "active" or "very active." This was most likely due to the fact that all data registries were obtained from an app update (MH version 2.0) which, naturally, received many of its users from the previous version. This could also partially explain why 62% of our users were on the Pro version (paid subscription). MH always offered a free training program upon first registration, so the newest users would be expected to be on a free deal, while more experienced users would naturally progress to payment modes.

The studied sample primarily pursued "muscle gain" or "weight loss" fitness goals. A remarkably small (21.4%) percentage trained for "antiaging" purposes. We initially interpreted this finding as a sign that individuals were focusing mainly on "appearance," but this would have to be further investigated, as muscle gain [23,24], as well as weight loss [25], are also known markers of improved health [25] and consequently better aging.

Attrition Rates and Perceived Difficulty

It is well-established that attrition rates in mobile apps are extremely high. Meyerowitz-Katz et al [26], in their 2020 meta-analysis, stated that up to 98% of people only use apps for a short period of time. Our results fully align with this marked tendency, as we appreciated a remarkable drop in the number of training sessions within the first few weeks of enrollment. There was an observable decline of 69.3% by the end of the first month, a reduction of 77.5% by the end of month 2, and an 80.6% decline by the end of month 3. These figures strike even harder if we assume that many enrollments allegedly came from MH users who were transitioning from the old to the new version of the app. Participants in our study preferred "afternoon" (12:31-20:00 hours) training sessions and declared mean rates of session "difficulty" of 5.56, over a total of 10 points. The "difficulty" variable and its results need to be interpreted with caution. In our study, "difficulty" was an equivalent of perceived exertion, and it aimed to be indicative of how hard the session had felt to the user. However, this data were inquired once the user had not only finished the training but also finished the cool-down phase, and this could have led to respondents underrating the perceived exertion derived from the main block of training. Contrary to our expectations, difficulty in our sample showed a strong positive correlation to retention, which could be interpreted as a sign that challenge fosters engagement. Indeed, there is previous evidence that complex, vigorous, or hybrid activities correlate with intrinsic motivation [27], which commonly underlines activity retention. Regarding constructs of adherence, this finding could also reflect a self-selection bias, where more committed users are more likely to opt for challenging sessions, thus reinforcing their engagement over time.

Factors Influencing Training Frequency

Frequency of training seemed not to be related to factors such as age, BMI, declared enjoyment or perceived difficulty, subscription vs nonsubscription, declared fitness goal, or preferred training schedule. However, statistically significant differences were observed based on sex, previous activity level, motivation, and body type. Frequency of training was greater in the introjected motivation group, in males, in the very active vs inactive, and in the thin vs mid groups. One could argue that the controlled and external regulation of introjected motivation could explain the increased frequency observed in this group. This would partially align with previous research that points to the primacy of extrinsic motivation in exercise contexts [28]. The fact that introjected motivation seems to encourage higher training frequencies but no longer retention or higher adherence might be indicative of an enthusiasm that is not sustained over time. As to the user's previous activity level, while it seems logical that highly active individuals would train more often, this could be influenced by their prior engagement with the MH app or other forms of PA. If they were former MH users, their higher frequency could indicate loyalty, whereas if their activity stemmed from external sources, it is noteworthy that they also engaged frequently with the app. In contrast, inactive individuals may have felt overwhelmed by structured training. Previous research highlights differences in how beginners perceive social comparison and networking features in fitness apps, as well as how exercise proficiency affects adherence [29]. Additionally, attitudes toward PA significantly impact behavior, with Feng et al [30] showing that greater activity levels correspond to deeper integration and sustained engagement. Another possible explanation is that very active users may use more app features, enhancing their overall experience and leading to higher engagement [30]. Our results should, however, be interpreted with caution, since despite statistical significance, effect sizes

were small to very small, which is indicative of them having limited practical implications.

Reflections on Adherence to mHealth Training

In our study, adherence did not correlate with age, sex, previous level of activity, declared fitness goal, being on a free plan versus subscription mode, training schedule, perceived difficulty, or enjoyment in sessions.

Only one statistically significant difference was found for adherence, and it was for the "thin" group, which showed higher adherence than the "mid" group. Both frequency and adherence in this study were correlated with "thin" body type, but in both cases the effect size was small, so the association may not imply high practical impact. Notably, no motivation type proved to be more relevant for adherence, in spite of several authors having pointed to the more autonomous regulations of motivation leading to increased adherence and persistence [31]. Recent evidence confirms that maintaining physical activity remains challenging for healthy adults, with persistent individual-level barriers (ie, lack of motivation, attitudes, and concerns about physical changes) [32]. Adherence results in our study ranged from 1.2% to 166.7%, which was an impactfully wide range. It is important to note that intensity and duration data were not consistently available across users, which limited our ability to construct the adherence measure. Adherence in this study was based only on training frequency, a limitation that highlights the need for more standardized and comprehensive adherence metrics in future app-based exercise research. These results brought us to the following insights. Adherence is rather a measure of precise forecasting, as it basically depends on the ability to foresee future behavior. In that case, several personal characteristics may come into play, which have not been assessed in this study, such as the concept of self-efficacy, ambition, the ability to plan in advance, or the ability to pursue goals. Similarly, in healthy adults, psychological factors such as self-efficacy, enjoyment, and planning were significant predictors of long-term adherence to PA, emphasizing the relevance of individual motivational and behavioral traits in sustained engagement [33]. In addition, a study identified lack of time, motivation, and fatigue as frequent barriers to PA in healthy young adults, while enjoyment and social support emerged as consistent facilitators [34]. We found the lowest adherence rates for those who trained 6 times per week, while the highest adherence values were for those who trained twice per week. Based on our results, individuals with lower frequency expectations managed better to fulfill their target plan and were, consequently, more adherent. Again, we see a disadvantage in how different researchers seem to measure and define adherence, in addition to the fact that electronically delivered interventions often lack a detailed reporting of it [35]. We note that in our study, adherence was based on training frequency (% of targeted versus actual), which naturally correlates both variables. In contrast, retention and adherence operate on different parameters, especially when the exercise program is nonprescribed, lacks external obligation, and has no set duration.

Retention as a Key Variable, Distinct From Adherence

In this study, retention correlated significantly with most study variables. There was higher retention in the male group, in subscribers, in "active" and "very active," in "mid" body types versus "thin," and also higher retention when the fitness goal was "antiaging" or "gain muscle" versus "lose weight." "Pro version" users exhibited higher retention, aligning with previous research linking price to commitment [29,36,37], suggesting that subscription may indicate greater interest. The effect size for our finding was moderate-to-large, which points at subscription possibly being the most determinant factor in long-term training behavior. All other correlations had small to very small effect sizes. "Thin" body types, which had shown correlation with frequency and adherence, did not display higher retention, potentially reflecting an initial enthusiasm that wanes over time. The finding that "antiaging" goals led to higher retention than "lose weight" aligns with theories suggesting that health-oriented goals promote sustained engagement. However, the fact that "gain muscle" goals also outperformed "lose weight" in retention suggests that aesthetic-driven objectives may still play a role in long-term engagement, challenging this interpretation.

The study found that only intrinsic motivation had a statistically significant positive correlation with retention, while no such correlation was observed between intrinsic motivation and adherence. The distinction between retention and adherence is emphasized, as these concepts are considered distinct. In line with previous evidence [28,38-40], our study reflects that intrinsic and identified autonomously regulated motivations are the strongest correlated with retention. There is, however, previous evidence that points at extrinsic regulations of motivation as possibly the most important ones for exercise contexts [28,41]. We agree with Wilson's statement that future research with larger sample sizes is recommended, considering potential variations in extrinsic motivation types [28] and a revision of the commonly accepted theory that intrinsic motivation is the most desirable to engage in and sustain exercise activities.

To translate our findings into practical applications, we suggest that fitness app developers consider tailoring features to specific user subgroups (ie, providing targeted support or content adaptations for users with a "mid" body type). Moreover, including motivational aspects that support intrinsic regulation, such as goal-tracking tools, personalized feedback, and autonomy-enhancing design, may further increase retention and adherence.

Strengths and Limitations

The present study focused on analyzing user training behavior by means of a cross-sectional study conducted on 2771 MH app users over a period of 18 months. To the authors' knowledge, the largest study previously available was a cohort study conducted under the Konstanz Life Study, which followed a total of 1236 users of either fitness or nutrition apps [21]. Other revisions involved larger samples, such as that by Stecher et al [8], (with 3555 participants across 22 interventions) or He et al [7] (with 1622 participants from 6 studies). Previous studies followed participants for periods of 2 to 24 months [7,8,22].

Based on this evidence, our study could be the largest of its kind in sample size and follow-up period so far. Nonetheless, we acknowledge that broader meta-analyses may include larger cumulative samples and aggregated durations across multiple interventions and apps [26].

The use of real-world app data, combined with motivational surveys, provides valuable insights into user behavior, retention, and adherence patterns. Additionally, the study uses robust statistical analyses, including nonparametric tests and multiple correction methods, ensuring the reliability of the findings.

However, the research also has limitations. The cross-sectional design prevents establishing causal relationships between motivation, training behavior, and adherence. The dataset is limited to users of a single fitness app (MH), potentially restricting generalizability to other platforms with different features or user demographics. We obtained informed consent from 5858 users. Of those registries, 2771 were complete and eligible for analysis. This loss should be acknowledged as the fact that we only managed to merge motivation and training data for a total of 289 participants, which limits the statistical power of our motivation-related analyses. Finally, adherence was measured in terms of training frequency, which may not fully capture engagement in app-based fitness programs, highlighting the need for more nuanced adherence metrics in future research.

Conclusions

This study provides crucial insights into the exercise behavior and retention patterns of MH app users, highlighting key factors that influence user engagement. New insights are shared in regard to how motivation relates to training behavior with fitness apps. Clear differentiations are presented between adherence and retention, as conceptualized by the study authors. Fitness apps are a promising tool toward more active lifestyles, but we are yet lacking a sound understanding of related human behavior. Strategies such as gamification, goal-setting, or prompting are available to app developers to increase user engagement. However, longitudinal studies and mixed methods approaches are needed both to study causality and explore qualitative drivers to trainee behavior.

Acknowledgments

We would like to thank all users of MH app who consent to share their anonymized data for research purposes. All authors declared that they had insufficient funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Funding

This study was supported with research funds provided by Aristos Campus Mundus projects (ACM2022_25 and ACM2023_19) and Fundació Naccari-Ravà 2022. The funding bodies had no role in study design, data collection, data analysis, or manuscript preparation.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: BG-Z, JMA-M, JM, JJ-R
Data curation: OB-J
Formal analysis: AF-V, RP, JMA-M
Funding acquisition: JMA-M, BG-Z, JM
Investigation: AF-V
Methodology: AF-V, RP, BG-Z, MG-B
Project administration: JM
Supervision: JM, BG-Z, JJ-R
Validation: AF-V
Visualization: JM, JJ-R
Writing – original draft: AF-V, JM
Writing – review & editing: AF-V, RP, JMA-M, OB-J, MG-B, JM, JJ-R

Conflicts of Interest

None declared.

Multimedia Appendix 1

Complete description of the Mammoth Hunters dataset, including all variable categories and the full motivation survey used in the study.
[DOCX File , 20 KB - mhealth_v14i1e72201_app1.docx]

References

1. WHO guidelines on physical activity and sedentary behaviour: at a glance Internet. World Health Organization. Geneva PP - Geneva: World Health Organization URL: https://iris.who.int/handle/10665/337001 [accessed 2025-12-04]
2. Rodgers WM, Hall CR, Duncan LR, Pearson E, Milne M. Becoming a regular exerciser: examining change in behavioural regulations among exercise initiates. Psychology of Sport and Exercise 2010;11(5):378-386 [FREE Full text] [doi: 10.1016/j.psychsport.2010.04.007]
3. Bailey DL, Holden MA, Foster NE, Quicke JG, Haywood KL, Bishop A. Defining adherence to therapeutic exercise for musculoskeletal pain: a systematic review. Br J Sports Med 2020;54(6):326-331 [FREE Full text] [doi: 10.1136/bjsports-2017-098742] [Medline: 29875278]
4. Number of US Fitness Health App Users Internet. Statista. URL: https://www.statista.com/statistics/1154994/number-us-fitness-health-app-users/ [accessed 2025-12-04]
5. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
6. Yang Y, Boulton E, Todd C. Measurement of adherence to mHealth physical activity interventions and exploration of the factors that affect the adherence: scoping review and proposed framework. J Med Internet Res 2022;24(6):e30817 [FREE Full text] [doi: 10.2196/30817] [Medline: 35675111]
7. He Z, Hassan MA, Saiz-González P, Ryu S, Wang R, Gao Z. Smartphone app-based interventions on physical activity behaviors and psychological correlates in healthy young adults: a systematic review. PLoS One 2024;19(4):e0301088 [FREE Full text] [doi: 10.1371/journal.pone.0301088] [Medline: 38578729]
8. Stecher C, Pfisterer B, Harden SM, Epstein D, Hirschmann JM, Wunsch K, et al. Assessing the pragmatic nature of mobile health interventions promoting physical activity: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023;11:e43162 [FREE Full text] [doi: 10.2196/43162] [Medline: 37140972]
9. Ozdamli F, Milrich F. Positive and negative impacts of gamification on the fitness industry. Eur J Investig Health Psychol Educ 2023;13(8):1411-1422 [FREE Full text] [doi: 10.3390/ejihpe13080103] [Medline: 37623300]
10. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
11. Lu T, Lin Q, Yu B, Hu J. A systematic review of strategies in digital technologies for motivating adherence to chronic illness self-care. NPJ Health Syst 2025;2(1). [doi: 10.1038/s44401-025-00017-4]
12. Fang P, Shi S, Menhas R, Laar RA, Saeed MM. Demographic characteristics and digital platforms for physical activity among the Chinese residents during the COVID-19 pandemic: a mediating analysis. JMDH 2022;15:515-529. [doi: 10.2147/jmdh.s354984]
13. Al-Shamaileh O, Sutcliffe A. Why people choose apps: an evaluation of the ecology and user experience of mobile applications. Int J Hum Comput Stud 2023;170:102965. [doi: 10.1016/j.ijhcs.2022.102965]
14. Peng X, Menhas R, Dai J, Younas M. The COVID-19 Pandemic and overall wellbeing: mediating role of virtual reality fitness for physical-psychological health and physical activity. Psychol Res Behav Manag 2022;15:1741-1756. [doi: 10.2147/prbm.s369020]
15. Liu R, Menhas R, Dai J, Saqib ZA, Peng X. Fitness apps, live streaming workout classes, and virtual reality fitness for physical activity during the COVID-19 lockdown: an empirical study. Front Public Health 2022;10:852311 [FREE Full text] [doi: 10.3389/fpubh.2022.852311] [Medline: 35812515]
16. Yang J, Menhas R, Dai J, Younas T, Anwar U, Iqbal W, et al. Virtual reality fitness (VRF) for behavior management during the COVID-19 pandemic: a mediation analysis approach. Psychol Res Behav Mana 2022;Volume 15:171-182. [doi: 10.2147/prbm.s350666]
17. Luo C. An exploration of the impact of fitness apps on individual exercise behaviour. Rev Psicol del Deport 2024;33:385-393.
18. Boucher EM, Raiker JS. Engagement and retention in digital mental health interventions: a narrative review. BMC Digit Heal 2024;2(1):52. [doi: 10.1186/s44247-024-00105-9]
19. Deci EL, Ryan RM. Self-determination theory: a macrotheory of human motivation, development, and health. Can Psychol / Psychol Can 2008;49(3):182-185. [doi: 10.1037/a0012801]
20. Ryan RM, Deci EL. Intrinsic and extrinsic motivation from a self-determination theory perspective: definitions, theory, practices, and future directions. Contemp Educ Psychol 2020;61:101860. [doi: 10.1016/j.cedpsych.2020.101860]
21. König LM, Sproesser G, Schupp HT, Renner B. Describing the process of adopting nutrition and fitness apps: behavior stage model approach. JMIR Mhealth Uhealth 2018;6(3):e55 [FREE Full text] [doi: 10.2196/mhealth.8261] [Medline: 29535078]
22. Herrmann LK, Kim J. The fitness of apps: a theory-based examination of mobile fitness app usage over 5 months. Mhealth 2017;3:2 [FREE Full text] [doi: 10.21037/mhealth.2017.01.03] [Medline: 28293619]

23. Wang DXM, Yao J, Zirek Y, Reijnierse EM, Maier AB. Muscle mass, strength, and physical performance predicting activities of daily living: a meta-analysis. J Cachexia Sarcopenia Muscle 2020;11(1):3-25 [FREE Full text] [doi: 10.1002/jcsm.12502] [Medline: 31788969]
24. Westcott WL. Resistance training is medicine: effects of strength training on health. Curr Sports Med Rep 2012;11(4):209-216. [doi: 10.1249/jsr.0b013e31825dabb8]
25. Jayedi A, Khan TA, Aune D, Emadi A, Shab-Bidar S. Body fat and risk of all-cause mortality: a systematic review and dose-response meta-analysis of prospective cohort studies. Int J Obes (Lond) 2022;46(9):1573-1581. [doi: 10.1038/s41366-022-01165-5] [Medline: 35717418]
26. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
27. Hsu RMCS, Cardoso FL, Varella MAC, Pires EM, Valentova JV. Comparing different typologies of physical activities with a focus on motivation. Front Psychol 2022;13:790490 [FREE Full text] [doi: 10.3389/fpsyg.2022.790490] [Medline: 35645925]
28. Wilson PM, Rodgers WM, Fraser SN, Murray TC. Relationships between exercise regulations and motivational consequences in university students. Res Q Exerc Sport 2004;75(1):81-91. [doi: 10.1080/02701367.2004.10609136] [Medline: 15532364]
29. Hu J, He W, Zhang J, Song J. Examining the impacts of fitness app features on user well-being. Inf Manag 2023;60(5):103796 [FREE Full text] [doi: 10.1016/j.im.2023.103796]
30. Feng B. Exploration on how integration degree into sport APPs affect Users? Attitude and behavioural habit towards physical exercise. Rev Psicol del Deport 2024;33:14-23.
31. Rodrigues F, Bento T, Cid L, Pereira Neiva H, Teixeira D, Moutão J, et al. Can interpersonal behavior influence the persistence and adherence to physical exercise practice in adults? A systematic review. Front Psychol 2018;9:2141 [FREE Full text] [doi: 10.3389/fpsyg.2018.02141] [Medline: 30459690]
32. Gabay M, Oravitan M. The factors affecting adherence to physical activity in fitness facility settings: a narrative review. Timisoara Phys Educ Rehabil J 2022;15:46-61. [doi: 10.2478/tperj-2022-0013]
33. Watson-Mackie K, Arundell L, Lander N, McKay FH, Jerebine A, Venetsanou F, et al. Technology-supported physical activity and its potential as a tool to promote young women's physical activity and physical literacy: systematic review. J Med Internet Res 2024;26:e52302 [FREE Full text] [doi: 10.2196/52302] [Medline: 39423006]
34. Duffey K, Barbosa A, Whiting S, Mendes R, Yordi Aguirre I, Tcymbal A, et al. Barriers and facilitators of physical activity participation in adolescent girls: a systematic review of systematic reviews. Front Public Health 2021;9:743935 [FREE Full text] [doi: 10.3389/fpubh.2021.743935] [Medline: 34722450]
35. Valenzuela T, Okubo Y, Woodbury A, Lord SR, Delbaere K. Adherence to technology-based exercise programs in older adults: a systematic review. J Geriatr Phys Ther 2018;41:49-61. [doi: 10.1519/jpt.0000000000000095]
36. Oyibo K. Investigating the key persuasive features for fitness app design and extending the persuasive system design model: a qualitative approach. Proc Int Symp Hum Factors Ergon Healthc 2021;10(1):47-53 [FREE Full text] [doi: 10.1177/2327857921101022]
37. James TL, Bélanger F, Lowry PB. The mediating role of fitness technology enablement of psychological need satisfaction and frustration on the relationship between goals for fitness technology use and use outcomes. J Assoc Inf Syst Internet 2022;23:913-965 [FREE Full text] [doi: 10.17705/1jais.00745]
38. Kuroda Y, Sato Y, Ishizaka Y, Yamakado M, Yamaguchi N. Exercise motivation, self-efficacy, and enjoyment as indicators of adult exercise behavior among the transtheoretical model stages. Glob Health Promot 2012;19(1):14-22. [doi: 10.1177/1757975911423073] [Medline: 24801311]
39. Ryan RM, Deci EL. Overview of Self-Determination theory: An Organismic-Dialectical Perspective. Handb self-Determination Res. Rochester, NY, US: University of Rochester Press; 2002:3-33.
40. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
41. Ryan RM. Psychological needs and the facilitation of integrative processes. J Pers 1995;63(3):397-427. [doi: 10.1111/j.1467-6494.1995.tb00501.x] [Medline: 7562360]

Abbreviations

MH: Mammoth Hunters
mHealth: mobile health
PA: physical activity
SDT: self-determination theory
WHO: World Health Organization

Edited by L Buis; submitted 13.Feb.2025; peer-reviewed by RN Al-Jabali, R Menhas; comments to author 11.Jun.2025; revised version received 23.Jul.2025; accepted 06.Nov.2025; published 08.Jan.2026.

Please cite as:
Fuente-Vidal A, Prat R, Arribas-Marin JM, Bastidas-Jossa O, Guerra-Balic M, Garcia-Zapirain B, Montane J, Jerez-Roig J
Analysis of Training Behavior in Users of a Fitness App: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e72201
URL: https://mhealth.jmir.org/2026/1/e72201
doi:10.2196/72201
PMID:

Original Paper

Effectiveness of a Hybrid Community-Based Heart-Healthy Lifestyle Intervention: Three-Arm Randomized Controlled Trial Integrating mHealth and Motivational Interviewing

Jina Choo[1,2], DrPH, PhD; Yura Shin[1,2], MSN; Songwhi Noh[1,2], MSN; Juneyoung Lee[3], PhD

[1]College of Nursing, Korea University, Seoul, Republic of Korea

[2]Transdisciplinary Major in Learning Health Systems, Graduate School, Korea University, Seoul, Republic of Korea

[3]Department of Biostatistics, College of Medicine, Korea University, Seoul, Seoul, Republic of Korea

Corresponding Author:
Jina Choo, DrPH, PhD
College of Nursing
Korea University
145, Anam-ro, Seongbuk-gu
Seoul, 02841
Republic of Korea
Phone: 82 232904925
Email: jinachoo@korea.ac.kr

Abstract

Background: Limited empirical evidence exists on the effectiveness of a hybrid approach to heart-healthy lifestyle interventions that integrates mobile health (mHealth) technology with face-to-face counseling. Moreover, its superiority over exclusive mHealth use in promoting heart-healthy behavioral outcomes within a community setting remains unclear.

Objective: This study aims to evaluate the effectiveness of a hybrid community-based approach to heart-healthy lifestyle intervention incorporating a mobile app and motivational interviewing among community-dwelling adults without a history of cardiovascular disease.

Methods: We conducted a 3-arm, parallel-group, randomized controlled trial with assessments at baseline and after 12 weeks. A total of 75 participants, each presenting at least 1 component of metabolic syndrome and no history of cardiovascular disease, were randomly assigned to 1 of 3 groups: hybrid (n=25), mobile (n=25), or control (n=25). Participants were recruited through an online platform. The hybrid group underwent a 12-week hybrid intervention combining a mobile app (ie, “My HeartHELP”) and face-to-face motivational interviewing led by a nursing researcher. The mobile group used only the mobile app, while the control group received written material on general heart health. The intervention was facilitated by 3 trained nursing researchers. The primary outcome was a composite score of “heart-healthy behaviors,” while secondary outcomes included scores for heart-healthy “information,” “self-efficacy,” “motivation,” and cardiovascular parameters. The trial was conducted in 2 rounds from October 2022 to May 2023. An intention-to-treat analysis was performed.

Results: Of the 75 participants, 72 (96%) completed this study. Compared with the control group, both the hybrid and mobile intervention groups demonstrated significantly greater improvements in behavioral outcomes, including composite heart-healthy behavior ($F_{2,69}$=7.25, P=.001), its theoretical predictors—heart-healthy motivation ($F_{2,69}$=8.54, P<.001) and self-efficacy for diet ($F_{2,69}$=4.87, P=.01) and exercise ($F_{2,69}$=5.48, P=.006)—as well as fasting glucose levels ($F_{2,69}$=3.90, P=.03) following the 12-week intervention. Particularly, the hybrid group—unlike the mobile group—showed significantly greater improvement in dietary behavior, a subscale of heart-healthy behavior, compared with the control group, and demonstrated significantly greater improvements in interest or enjoyment, a core subscale of intrinsic motivation, than the mobile and control groups.

Conclusions: The hybrid community-based heart-healthy lifestyle intervention—integrating a mobile app and motivational interviewing—demonstrated overall effectiveness comparable to the mobile app alone, while yielding greater improvements in dietary behavior and core intrinsic motivation. These findings highlight the potential of mHealth apps as practical, stand-alone tools to promote cardiovascular health, particularly in community settings with limited access to in-person professional support. However, incorporating motivational interviewing may further enhance internalized motivation and complex behavior changes over time. Health professionals can therefore adopt mHealth either independently or in combination with motivational interviewing.

Future studies should optimize integration strategies to enhance effectiveness and evaluate the long-term sustainability of such hybrid approaches.

Trial Registration: ISRCTN Registry ISRCTN83643383; https://www.isrctn.com/ISRCTN83643383

International Registered Report Identifier (IRRID): RR2-10.1161/circ.147.suppl_1.P147

(JMIR Mhealth Uhealth 2026;14:e76521) doi:10.2196/76521

KEYWORDS

healthy lifestyle; cardiovascular diseases; primary prevention; mobile applications; motivational interviewing; mobile phone

Introduction

Cardiovascular disease remains the leading cause of mortality worldwide [1]. To mitigate cardiovascular disease risk, the World Health Organization advocates for lifestyle modification interventions targeting smoking, poor dietary habits, physical inactivity, obesity, and alcohol consumption [2]. Evidence indicates that multiple unhealthy lifestyle behaviors significantly elevate cardiovascular disease risk [3], whereas adopting multiple healthy lifestyle behaviors can reduce all-cause mortality by up to 58%-66% [3,4]. Despite these findings, only a small proportion of individuals at risk of cardiovascular disease in the United States have achieved these recommended lifestyle changes [5]. Consequently, there is a pressing need for innovative, integrated strategies to effectively promote heart-healthy lifestyles in a comprehensive approach.

To date, mobile health (mHealth) interventions using websites, smartphones, or videoconferencing (online) have been used to improve heart health outcomes, such as cardiovascular risk parameters [6]. mHealth may help offset limited professional human resources and the time and effort required for expert-centered cardiovascular health care in the community [7]. However, evidence on the effectiveness of mHealth interventions in changing behaviors remains limited [8]. When studied, they have primarily focused on a single approach, such as physical activity [9] or weight loss [10], rather than a comprehensive strategy targeting multiple lifestyle behaviors.

Furthermore, mHealth interventions may face challenges in maximizing educational effects due to a lack of empathic support from the absence of active human interaction [11]. The primary limitation of mHealth interventions may be characterized as the lack of a robust therapeutic alliance, as highlighted by Müssener [12]. To overcome these limitations, mHealth interventions may incorporate tailored interactions using real-time feedback messages as an innovative strategy to strengthen therapeutic alliance and promote heart-healthy behavioral changes [9]. For this purpose, mHealth interventions require technologies that integrate theory-based cognitive behavioral strategies, such as goal-setting, self-monitoring, and reinforcement or feedback [13]. Based on this background, we developed a mobile app called “My HeartHELP,” incorporating 3 behavioral strategies to promote comprehensive heart-healthy behaviors: providing information, encouraging self-monitoring, and delivering real-time automatized feedback messages [14]. The behavioral strategies of “My HeartHELP” may be innovative in the ability to address multiple heart-healthy behaviors simultaneously while tailoring real-time feedback to individual behavioral outcomes. Hence, evidence on the effectiveness of the “My HeartHELP” app on heart-healthy behavioral outcomes is needed.

Nurse-led face-to-face behavioral lifestyle interventions are still required, as they elicit small-to-moderate effects on behavioral changes [15]. However, their effectiveness in modifying lifestyle behaviors remains unclear in community settings without a system that educates and supports community-dwelling individuals in self-management. Among the several behavioral intervention modes proposed, motivational interviewing has been recognized as a potential approach to facilitate behavioral changes through person-centered care, fostering active interaction between the client and provider within the client’s psychosocial context [13]. This method involves counseling techniques such as reflections, affirmations, open questions, and summarizations, which are well-received by clients because they address internal conflicts and guide goal-setting [16]. By focusing on connecting behaviors to desired outcomes and evoking an individual’s intrinsic motivation, motivational interviewing encourages people to take an active role in their own change processes [17]. In the context of behavioral interventions, motivational interviewing has proven effective in managing substance use, smoking cessation, and physical activity [18]. Therefore, motivational interviewing may complement mHealth by addressing its shortcomings and could be integrated into mHealth interventions in community settings.

A hybrid approach integrating mHealth with face-to-face motivational interviewing counseling presents a promising alternative for addressing the limitations of mHealth. This approach offers 2 key advantages. First, it combines the technological convenience of mHealth with the person-centered care of face-to-face interventions, thereby complementing the constraints of online interventions in terms of therapeutic alliance [19]. Second, in-person interactions can enhance engagement in and adherence to lifestyle behaviors recommended by health care expert providers [20]. In this regard, the American Heart Association has posited that a combined approach involving multiple modalities of behavioral interventions targeting cardiovascular risk factor reduction may be more effective than a single-modality approach [13]. Furthermore, the addition of clinical expert counseling to mobile-based cardiovascular health programs has not only been recommended [21] but has also been associated with greater reductions in systolic blood pressure when nurse-led case management was integrated with mHealth interventions, compared to mHealth alone [22]. However, empirical research on whether hybrid interventions that combine online mHealth

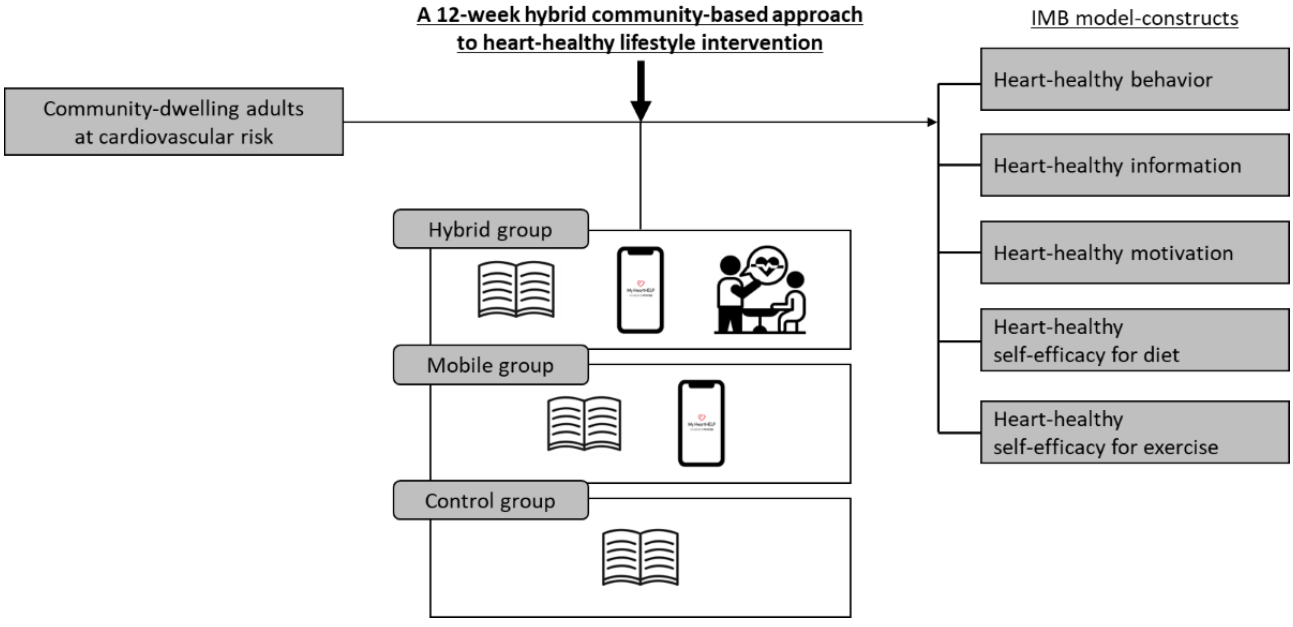
and offline face-to-face interventions are more effective than mHealth alone remains limited.

To address the current gap in the literature, we developed and implemented a hybrid community-based approach to a heart-healthy behavioral lifestyle intervention that combines the use of the "My HeartHELP" mobile app with face-to-face motivational interviewing to maximize behavioral outcomes in cardiovascular prevention and promotion. However, little information exists on whether such a hybrid approach is effective and, furthermore, whether it is superior to mobile app use alone in achieving heart-healthy behavioral outcomes within a community setting.

Meanwhile, the Information-Motivation-Behavioral Skills (IMB) model (Multimedia Appendix 1) served as the basis for developing the behavioral strategies of the hybrid community-based approach to a heart-healthy behavioral lifestyle intervention and determining the outcome variables to evaluate its effectiveness. The IMB model explains behavioral changes using constructs such as information, motivation, and behavioral skills [23]. Theoretically, motivation in the IMB model is directly associated with behavioral skills (ie, self-efficacy) and indirectly associated with behavior via self-efficacy [23]. Motivation enhancement may effectively change heart-healthy behaviors when used as part of lifestyle interventions [24]. In particular, the hybrid community-based approach was designed to enhance personal motivation by delivering information and behavioral skills for change through mobile text messaging and individually tailored counseling, while fostering self-monitoring to promote greater self-awareness and adoption of heart-healthy behaviors. Furthermore, its group sessions—grounded in motivational interviewing—were intended to facilitate the exchange of behavioral strategies and experiences through group dynamics, thereby strengthening social support and shared norms and, in turn, enhancing social motivation.

Therefore, the hybrid intervention based on the IMB model has a theoretical foundation aimed at informing individuals about cardiovascular health through frequent text messaging (heart-healthy information) and encouraging them through mHealth and motivational interviewing to promote intrinsic motivation (heart-healthy motivation). This, in turn, enhances behavioral skills (heart-healthy self-efficacy for diet and exercise) through cognitive-behavioral strategies and facilitates behavioral changes (heart-healthy behavior; Figure 1).

Figure 1. A theoretical framework of this study. The hybrid group received a 12-week hybrid intervention combining a mobile app (ie, "My HeartHELP") and face-to-face motivational interviewing; the mobile group used the mobile app (My HeartHELP), and the control group received a brochure on general heart-health. IMB: Information-Motivation-Behavior Skills.



We aimed to examine the effectiveness of a 12-week hybrid community-based heart-healthy lifestyle intervention among community-dwelling adults without a history of cardiovascular disease (Figure 1). We tested 2 hypotheses: first, a mobile group would be more likely to improve heart-healthy information, motivation, behavioral skills, and behavior (ie, IMB model-constructs) and cardiovascular parameters than a control group. Second, a hybrid group would be more likely to improve these outcomes than the mobile or control groups.

Methods

Study Design

This study (clinical trial number ISRCTN83643383) was a randomized controlled trial with 3 arms, parallel groups, and a 12-week follow-up period: a hybrid group, a mobile app group, and a control group. The pretest (T1) was conducted before the 12-week intervention, while the posttest (T2) was conducted after the intervention. This study was conducted in 2 cohort rounds to facilitate the effective implementation of the intervention by optimizing resources within a small research team. The rationale for the separate rounds was based on the

study by Burke et al [25]. The first round was conducted between November 2022 and February 2023, and the second round between February and May 2023 in a community-based setting in Seoul, South Korea.

As this study adopted a prospective, randomized, open-label, blinded end point design [26], double blinding was not feasible due to the nature of the interventions, which involved mobile app use and counseling. Specifically, neither study participants nor interviewers were blinded to this study's groups, but outcome assessors were blinded. We report our trial using the CONSORT (Consolidated Standards of Reporting Trials) guidelines (checklist provided in Multimedia Appendix 2) [27].

Participants

This study's participants comprised 75 community-dwelling adults, and the inclusion criteria were as follows: (1) aged between 20 and 64 years, and (2) having at least one metabolic syndrome according to the National Cholesterol Education Program Expert Panel [28]. The exclusion criteria were as follows: (1) medically diagnosed with diabetes mellitus and taking hypoglycemics or insulin, (2) medically diagnosed with either cardiovascular or psychiatric disease (ie, major depression or anxiety disorder), (3) physical activity limitations, (4) cognitive problems, and (5) inability to use mobile apps.

Participants were recruited using a notice on the internal bulletin board of the university with which the researchers were affiliated, as well as on the external bulletin boards of mobile communication platforms. The first round of recruitment was conducted from October 17 to November 11, 2022, and the second round from January 11 to February 3, 2023. When interested candidates contacted the researchers by phone to participate in this study, the researchers determined whether they were eligible and informed them that they could be randomly assigned to 3 groups. After agreeing to participate, participants were asked to attend a preliminary investigation session in person, where the purpose, methods, and procedures of this study were explained in detail based on this study's protocol, and written informed consent was obtained.

Sample Size

The minimum sample size per group was calculated based on the effect size of the composite heart-healthy behavior—the primary outcome variable in this study—between the mobile and control groups. According to Park [29], the mean difference (MD; Δ=change in mean composite heart-healthy behavior score from pretest to posttest) was 0.66 and –0.15 in intervention and control groups, respectively. However, we conservatively assumed that there would be no difference in the composite heart-healthy behavior between the pre- and posttests (ie, 0.0) in the control group. Thus, the effect size assumed in this study was 0.66.

The SD of the mean score changes for each group was estimated as follows, based on the results from the dissertation by Park [29]: pretest score SDs were assumed to be 0.47 and 0.56, and the posttest score SDs were 0.41 and 0.50 for the intervention and control groups, respectively. By conservatively assuming a correlation coefficient of 0.5 between pre- and posttests, and using the method proposed by Abrams et al [30], we projected the SD of the mean score changes from pre- to posttests to be 0.443 and 0.533 for the intervention and control groups, respectively. With these SDs of mean score changes and, under a 2-sided significance level (α) of 2.5%, a minimum of 20 subjects per group was needed to ensure a statistical power (1-β) of 95% to detect the effect size of 0.66. Assuming a 20% dropout rate during this study's period, we planned to enroll 25 participants in each group. Statistical software PASS 2020 (version. 20.0.8; NCSS LLC) was used to calculate the sample size.

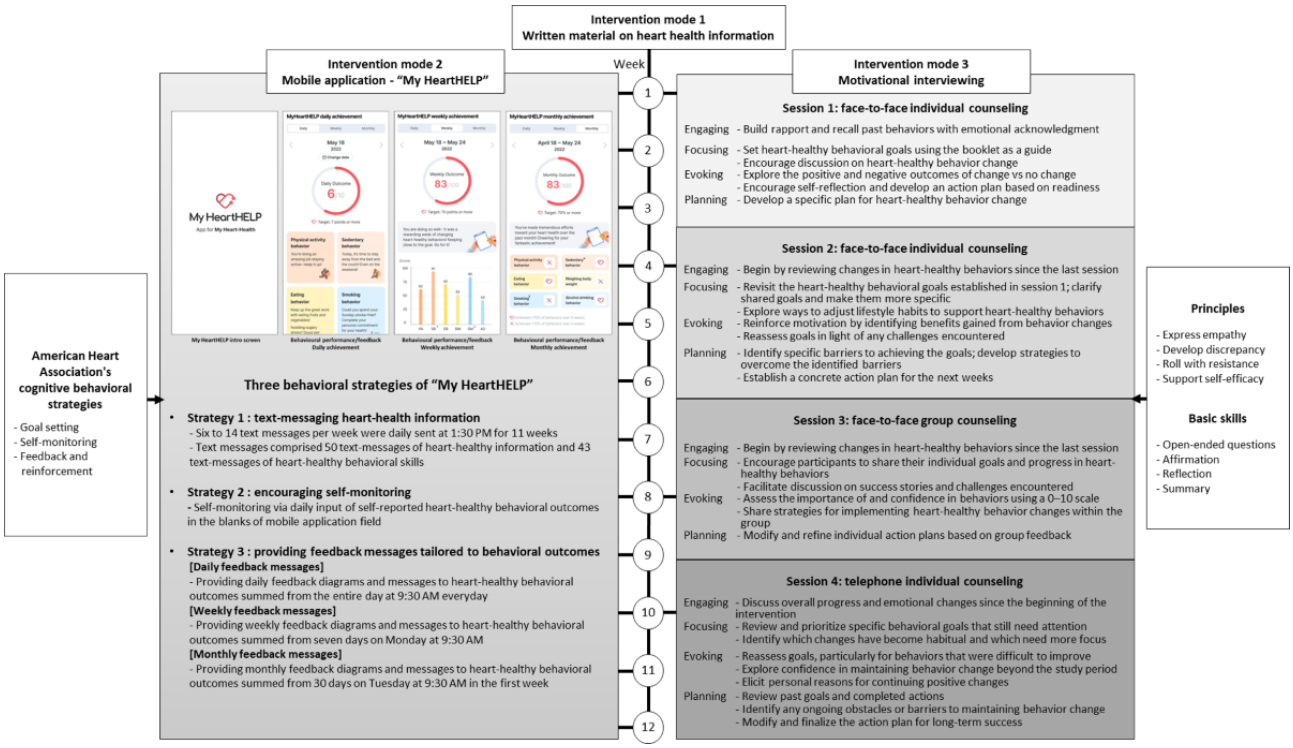
Random Allocation

Random allocation in this study was conducted by the principal investigator using an age- and gender-stratified block randomization method with an allocation ratio of 1:1:1 across the 3 study arms. A randomization list was generated for each of the 10 age- and gender-stratified groups (ie, 20-29, 30-39, 40-49, 50-59, and 60-65 years, by gender) using R software (version 3.6.0; R Foundation for Statistical Computing). Study participants were listed in the order in which they were recruited, and each participant was assigned to an appropriately stratified group based on age and sex. Finally, a total of 75 participants were allocated to the hybrid (n=25), mobile (n=25), or control (n=25) groups.

Study Intervention

An outline and details of the hybrid intervention are presented in Figure 2, which contains three intervention modes: (1) intervention mode 1: written materials on general heart-health information, (2) intervention mode 2: the "My HeartHELP" mobile app [14], and (3) intervention mode 3: motivational interviewing counseling. General heart health information was delivered via a brochure, including information on cardiovascular diseases, cardiovascular risk factors, and general heart-healthy lifestyle changes. The duration of intervention modes 2 and 3 was 12 weeks. Specifically, the mobile app was designed for daily use throughout the intervention period, while motivational interviewing was delivered in 4 sessions, each lasting 2 hours, conducted at weeks 1, 4, 8, and 10.

Figure 2. An outline for the 12-week hybrid community-based heart-healthy lifestyle intervention. The 12-week hybrid intervention consisted of the three intervention modes: (1) intervention mode 1: written materials on general heart-health information, (2) intervention mode 2: the "My HeartHELP" mobile app, and (3) intervention mode 3: motivational interviewing counseling. General heart health information was delivered via a brochure, including information on cardiovascular diseases, cardiovascular risk factors, and general heart-healthy lifestyle changes. Please see Multimedia Appendix 3 for a larger version of this figure.



The hybrid group participated in all 3 intervention modes. The mobile group received 2 intervention modes of written materials and the "My HeartHELP" app. The control group received only the written materials.

Participants in both the hybrid and mobile groups attended separate 1-hour preintervention education sessions, which provided an overview of the 12-week intervention schedule and instructions on using the "My HeartHELP" app. A printed booklet containing a QR code for app download was distributed, and participants were informed that the app was freely available for research use. Each session also included a hands-on practice session, during which the research team assisted participants with app registration and guided them in entering sample data into the app.

The "My HeartHELP" app was implemented as an online mode, consisting of 3 strategies: strategy 1—text messaging with heart-health information, strategy 2—encouraging the self-monitoring of 6 lifestyle behaviors (physical activity, nonsedentary behavior, healthy dietary behaviors, nonsmoking, nonalcohol binge drinking, and daily weighing of body weight) [14], and strategy 3—providing automated or tailored feedback messages tailored to individual behavioral outcomes from self-monitoring (Figure 2). The text messaging intervention was delivered daily at 1:30 PM over a period of 1 to 11 weeks and comprised 50 messages with heart-health information and 43 messages focused on heart-healthy behavioral skills. Encouraging self-monitoring was also a daily practice, requiring participants to record behavioral outcomes for the 6 behaviors by entering data into the designated input fields of the app. Automated or tailored feedback messaging, including diagrams and texts, was daily, weekly, and monthly based. As a preparatory step toward developing automated and tailored feedback messaging, the research team established behavioral target goals and their ranges for daily, weekly, and monthly outcomes based on 6 heart-healthy behaviors. According to the ranges, a pool of feedback text messages was developed and embedded into the app's algorithm. For daily outcomes, scores obtained from self-monitoring of the 6 heart-healthy behaviors during the previous day were categorized as either >7 or <7, with the daily goal set at >7 out of a possible 10 points. The daily feedback messages were presented as a diagram and texts at 9:30 AM every day. For weekly outcomes, the average scores over the previous 7 consecutive days were converted to a 100-point scale and classified into 3 categories: ≥70, 50-69, and <50. The weekly goal was set at ≥70. The weekly feedback messages were delivered on Monday at 9:30 AM. For monthly outcomes, the mean scores across the preceding 4 weeks were dichotomized into ≥70 and <70. The monthly feedback messages were delivered as diagrams and text messages on Tuesday at 9:30 AM in the first week. The app automatically delivered tailored feedback messages through an embedded algorithm that operated according to predefined behavioral target goals and their corresponding ranges. Further details are available in the feasibility and acceptability study of the "My HeartHELP" mobile app [14].

Motivational interviewing counseling was an offline face-to-face intervention conducted in 4 sessions: sessions 1 (week 1) and 2 (week 4) for individualized counseling, session 3 (week 8) for group counseling, and session 4 (week 10) for telephone counseling (Figure 2). Each session followed the core processes

of engaging, focusing, evoking, and planning, as outlined in Figure 2 [17]. Engaging focused on building trust, understanding participants' unique perspectives, reviewing past behaviors, and evaluating emotional and behavioral changes [17]. Focusing aimed to clarify shared goals, making them more specific and actionable, while evoking strengthened participants' motivations for change by imagining potential outcomes and assessing their importance and confidence. Planning facilitated the development of practical strategies by identifying barriers to behavior change and determining ways to overcome them. The intervention was tailored to individual lifestyle contexts and used the principles of facilitating empathy, discrepancy, resistance, and self-efficacy along with open-ended questions, affirmations, reflections, and summaries techniques. Additional details are presented in Figure 2.

Fidelity of Intervention

To ensure the fidelity of the "My HeartHELP" app implementation, 2 measures used in previous studies [14,31] were applied: access and self-monitoring rates. Participants in both the hybrid and mobile groups were instructed to access the app daily and self-monitor their behavioral outcomes (ie, fill in the blanks of the behavioral components) daily for 12 weeks. Among those enrolled in the hybrid and mobile groups, 36/47 (76.6%) accessed the app at least once daily, while 45.7/47 (97.3%) engaged daily in self-monitoring of heart-healthy behaviors throughout the 12 weeks.

To ensure the fidelity of motivational interviewing, all authors completed training and certification through motivational interviewing courses, including 10 hours of fundamental information and 8 hours focused on health care [32]. One author implemented the motivational interviewing sessions over 12 weeks as an interventionist, with assistance from another author. To maintain counseling quality, an activity sheet template was developed and used for each session. Regarding attendance adherence, participants in the hybrid group attended 23/23 (100%) of the sessions.

Measures

Heart-Healthy Behavioral Outcomes Variables

The primary outcome was the composite score of heart-healthy behaviors. Secondary outcomes included scores for heart-healthy knowledge, heart-healthy self-efficacy, heart-healthy motivation, and cardiovascular parameters. All outcomes were measured at baseline (T1) and after 12 weeks (T2) using self-reported questionnaires, anthropometric measurements, and blood sampling.

Composite heart-healthy behavior refers to the extent to which participants practiced heart-healthy behaviors for cardiovascular health. This was measured using the Management Behaviors of Metabolic Syndrome Evaluation Tool developed by Kang [33] for individuals with metabolic syndrome. This tool consists of 36 questions: 8 on "physical activity and weight control," 16 on "dietary habits," 3 on "drinking and smoking," 3 on "stress," 2 on "sleep and rest," and 4 on "health check-up and management." The average score was calculated by summing the scores for each item (rated from "never" to "always" on a 4-point Likert scale) and dividing it by the number of items. Higher scores indicated greater heart-healthy behavior. Cronbach α was 0.92 in the study by Kang [33] and 0.91 in this current study.

Heart-healthy information refers to the level of knowledge regarding cardiovascular disease prevention. This was measured using the Heart-Healthy Information Questionnaire developed by Choo et al [34]. This tool consists of 50 questions, each answered as "true," "false," or "don't know." Participants who answered correctly received 1 point. If participants answered incorrectly or did not know, they received 0 points. The total score ranged from 0 to 50 points. The Kuder-Richardson formula 20 was 0.85 in the study by Choo et al [34] and 0.81 in this current study.

Motivation refers to intrinsic motivation levels for practicing heart-healthy behaviors. It was measured using the Intrinsic Motivation Inventory developed by McAuley et al [35], which we modified for the present study. Based on the inventory guidance indicating that subscales may be selectively used according to the study's focus [36], we used 5 of the 7 subscales, excluding 2 subscales. The pressure or tension subscale was excluded because it is more relevant to stressful activities such as competitive sports [37,38] than to heart-healthy behaviors. The relatedness subscale was also excluded because its items capture relational closeness—such as perceived distance, trust, and friendship with specific individuals—which are less relevant to the autonomous, self-directed process of adopting heart-healthy lifestyle habits. Consequently, 5 of the original 7 subscales were used—interest or enjoyment (7 items), perceived competence (6 items), effort or importance (5 items), perceived choice (7 items), and value or usefulness (7 items)—yielding a total of 32 items. The English version of the modified Intrinsic Motivation Inventory was minimally adjusted to fit heart-healthy behaviors, then translated into Korean separately by 3 Korean nursing scholars and consolidated into a single Korean version. A native English speaker back-translated this version, and the final English version was reviewed and confirmed by the original translators. Participants rated items using a 7-point Likert scale ranging from "completely disagree" to "strongly agree," with responses summed and averaged for each participant. Cronbach α was 0.85 in the study by McAuley et al [35] and 0.88 in this current study.

Heart-healthy self-efficacy refers to participants' confidence in their ability to engage in heart-healthy eating and exercise. It was measured using 2 instruments: the Self-Efficacy for Diet and Self-Efficacy for Exercise tools developed by Sallis et al [39]. The Self-Efficacy for Diet instrument comprises 20 items, while the Self-Efficacy for Exercise instrument has 12 items. Scores for eating and exercise habits were calculated by summing the responses on a 5-point Likert scale, ranging from "I cannot do this at all" to "I can definitely do this." Higher scores indicate higher self-efficacy levels. Cronbach α for eating was 0.84 in the study by Shin and Lach [40] and 0.91 in this current study. Cronbach α for exercise was 0.90 in the study by Sallis et al [39] and 0.93 in this current study.

Cardiovascular Parameters

Cardiovascular parameters comprised BMI, waist circumference, blood pressure, fasting glucose, total cholesterol,

low-density lipoprotein cholesterol, and high-density lipoprotein cholesterol. Body weight was measured after an overnight fast using the InBody270 scale (InBody Co, Ltd). Before weighing, participants wore a study-provided gown and removed their shoes. Height was measured using a wall-mounted stadiometer. BMI was calculated as weight (kg) divided by height (m^2). Waist circumference (cm) was measured twice using a Gullick II measuring tape at the midpoint between the lowest rib and the iliac crest, with the average of the 2 measurements used. Blood samples were drawn from the antecubital vein in the morning after a 10-hour overnight fast, without participants taking any current medications, including antihypertensive or lipid-lowering drugs. The samples were analyzed as described in a previous study [41].

Statistical Analysis

Participants' demographic and baseline characteristics were summarized using descriptive statistics, including frequency, percentage, mean, and SD. To test the homogeneity of the 3 groups' baseline measures, chi-square tests were used for categorical variables, while a 1-way ANOVA was performed for continuous variables (ie, age and study variables).

Both intention-to-treat and per-protocol approaches were used. The intention-to-treat analysis included all randomized participants regardless of adherence to the intervention, whereas the per-protocol population consisted of those who strictly adhered to the intervention protocol [42]. Missing values, which occurred due to nonresponse to the measurements, were not imputed but analyzed as observed. Analysis of covariance was used to compare outcome variables among the hybrid, mobile, and control groups at T2. In this analysis, T2 scores served as the dependent variable, study groups as the main exposure variables, and baseline scores for each respective outcome variable as confounders to adjust for individual differences at the start of the intervention. As a post hoc analysis, the Tukey least significant difference multiple comparison test was performed to identify significant differences between study groups in outcome variable changes. The analyses further examined significant group differences in the subscales of composite heart-healthy behavior and motivation. All statistical analyses were conducted using SPSS/WIN (version 28.0; SPSS Inc), with a 2-sided $P<.05$ considered statistically significant.

Ethical Considerations

This study was approved by the Institutional Review Board of Korea University (KUIRB-2022-0287-01). All participants provided written informed consent. All procedures were performed in accordance with the ethical standards of the Institutional Research Committee and the 2013 Declaration of Helsinki [43].

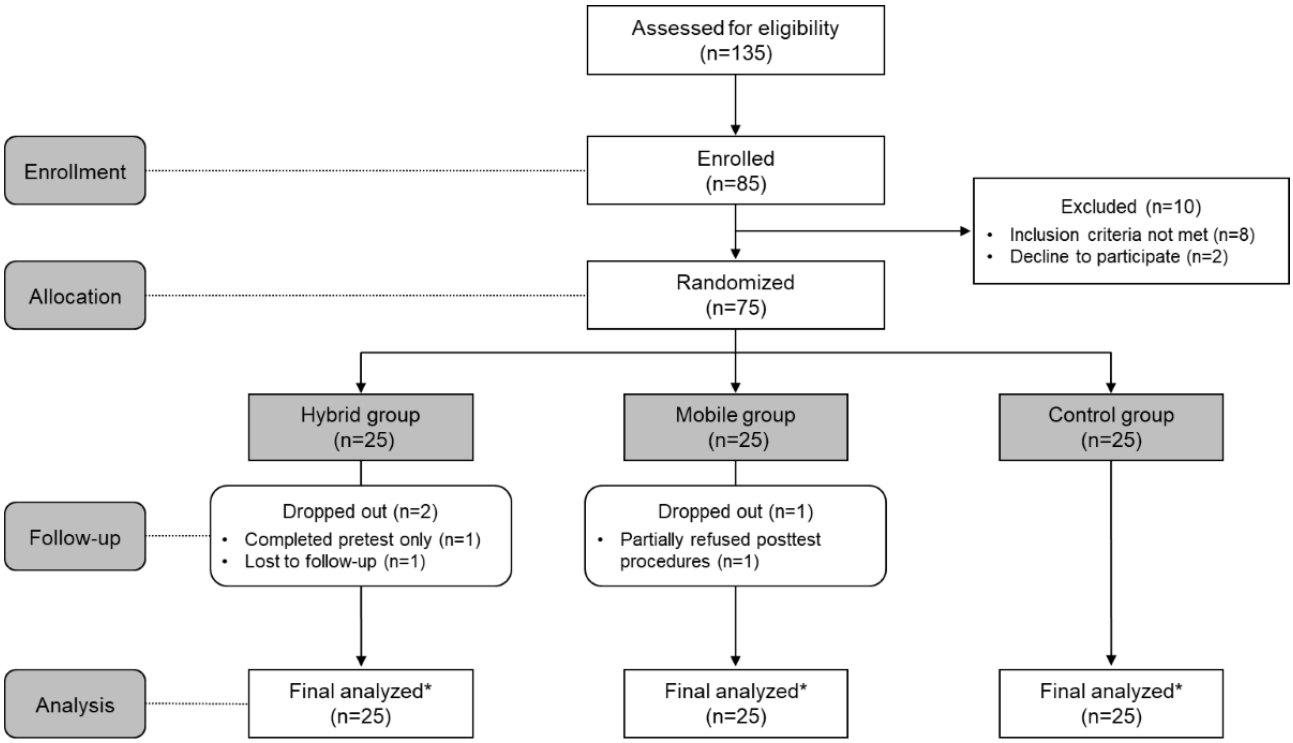
Safety and security procedures related to the use of the mobile app were carefully implemented. To ensure data privacy, access to the secure data server was restricted to a single designated member of the research team. Data were downloaded and processed offline to enhance data protection and confidentiality. Preintervention education sessions were conducted to inform participants about relevant safety and privacy measures. In addition, real-time user support was offered via KakaoTalk (Kakao Corp), a widely used mobile messaging platform in South Korea, enabling participants to directly communicate with the research team throughout the intervention period. Participants in the control group, mobile group, and hybrid group who completed only the pretest received a KR ₩30,000 (US $21) gift voucher. Participants who successfully completed all stages of the study received monetary compensation according to group assignment: KR ₩150,000 (US $104) for the control group, KR ₩250,000 (US $172) for the mobile group, and KR ₩300,000 (US $207) for the hybrid group. All details related to compensation were fully disclosed to participants during the informed consent process before their enrollment in this study.

Results

Participants' General Characteristics

Three participants dropped out of this study for the following reasons: one participant from the hybrid group due to time constraints (n=1), another from the hybrid group due to loss to follow-up (n=1), and one from the mobile group who declined to participate in the posttest (n=1; Figure 3). Since the findings from both the intention-to-treat and per-protocol analyses yielded identical statistical significance, the results are presented based on the intention-to-treat analysis.

Figure 3. Participant flow in this study. The hybrid group received a 12-week hybrid intervention combining a mobile app (ie, "My HeartHELP") and face-to-face motivational interviewing; the mobile group used the mobile app, and the control group received written material on general heart-health information. *All participants were included in an intention-to-treat analysis.



The participants had a mean age of 43.6 (SD 11.3) years (Table 1). Of all participants, 46 out of 75 (61.3%) were women, and 64 out of 75 (85.3%) were college-educated. Additionally, 32 out of 75 (42.7%) had a median income level greater than or equal to that of the general South Korean population [44]. Among the participants, 52 out of 75 (69.3%) were used, and 16 out of 75 (21.3%) were taking antihypertensive or lipid-lowering medications. No significant differences in general characteristics were observed between the groups (Table 1).

Table 1. Participants' general characteristics (N=75) those of the general South Korean population [44]. Among the participants, 52/75 (69.3%) were employed, and 16/75 (21.3%) were taking antihypertensive or lipid-lowering medications. No significant differences in general characteristics were observed between the groups.

	All (N=75)	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	Chi-square (*df*)	*F* test (*df*)	*P* value
Age (years), mean (SD)	43.6 (11.3)	43.4 (11.0)	43.9 (12.3)	43.6 (11.0)		0.01 (2, 69)	.99
Gender, n (%)					0.11 (2, 69)		.94
Women	46 (61.3)	16 (64)	15 (60)	15 (60)			
Men	29 (38.7)	9 (36)	10 (40)	10 (40)			
Education, n (%)					0.21 (2, 69)		.90
More than college-educated	64 (85.3)	22 (88)	21 (84)	21 (84)			
Less than college-educated	11 (14.7)	3 (12)	4 (16)	4 (16)			
Monthly household income[a], 10,000 won, n (%)					2.07 (2, 69)		.36
>500	32 (42.7)	11 (44)	8 (32)	13 (52)			
<500	43 (57.3)	14 (56)	17 (68)	12 (48)			
Employed, n (%)					1.63 (2, 69)		.44
Yes	52 (69.3)	15 (60)	18 (72)	19 (76)			
No	23 (30.7)	10 (40)	7 (28)	6 (24)			
Medications[b], n (%)					1.11 (2, 69)		.57
Yes	16 (21.3)	5 (20)	4 (16)	7 (28)			
No	59 (78.7)	20 (80)	21 (84)	18 (72)			

[a]Monthly household income was classified using a cutoff of 5,000,000 won (US $3450; IQR 1767-5300), which represents the median income of the general population in South Korea [44].

[b]Taking either antihypertensives or lowering lipids medications.

Heart-Healthy Behavioral Outcome Variables

Table 2 summarizes the results of the differential effects of the 3 groups on the primary and secondary outcome variables over a 12-week intervention period. Significant differences were observed among the 3 groups in the scores for composite heart-healthy behavior ($F_{2,69}$=7.25, *P*=.001), motivation ($F_{2,69}$=8.54, *P*<.001), self-efficacy for diet ($F_{2,69}$=4.87, *P*=.01), and self-efficacy for exercise ($F_{2,69}$=5.48, *P*=.006). Compared to the control group, the hybrid and mobile groups demonstrated significantly greater increases in the scores for composite heart-healthy behavior (hybrid vs control: MD 0.37, SD 0.10, *P*<.001; mobile vs control: MD 0.26, SD 0.10; *P*=.01), motivation (hybrid vs control: MD 0.74, SD 0.18; *P*<.001; mobile vs control: MD 0.42, SD 0.18; *P*=.02), self-efficacy for diet (hybrid vs control: MD 8.44, SD 2.83; *P*=.004; mobile vs control: MD 6.29, SD 2.77; *P*=.03), and self-efficacy for exercise (hybrid vs control: MD 5.01, SD 2.27; *P*=.03>; mobile vs control: MD 7.23, SD 2.24; *P*=.002). There were no significant differences in the heart-healthy information scores among the 3 groups.

Table 2. Effectiveness of a hybrid community-based heart-healthy lifestyle intervention on heart-healthy behavioral outcome variables (N=75). T1 and T2 refer to the baseline time point and the time point after 12 weeks, respectively.

	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	*F* test (*df*)	*P* value
Composite heart-healthy behavior, mean (SD)					
Overall				7.25 (2, 69)	.001
T1	2.4 (0.46)	2.4 (0.47)	2.4 (0.42)		
T2	2.9 (0.46)[a]	2.8 (0.42)[a]	2.5 (0.50)[b]		
Physical activity and weight control				4.02 (2, 69)	.02
T1	1.9 (0.73)	2.2 (0.71)	2.0 (0.57)		
T2	2.5 (0.56)[a]	2.7 (0.75)[a]	2.2 (0.68)[b]		
Dietary habits				5.04 (2, 69)	.009
T1	2.3 (0.50)	2.3 (0.60)	2.3 (0.55)		
T2	2.9 (0.51)[a]	2.7 (0.58)[a,b]	2.5 (0.62)[b]		
Drinking and smoking				0.06 (2, 69)	.94
T1	3.3 (0.91)	3.2 (0.81)	3.1 (0.92)		
T2	3.3 (0.85)	3.3 (0.91)	3.2 (0.89)		
Stress				1.54 (2, 69)	.22
T1	3.2 (0.69)	3.0 (0.70)	2.9 (0.53)		
T2	3.4 (0.54)	3.2 (0.66)	3.0 (0.63)		
Sleep and rest				1.88 (2, 69)	.16
T1	2.9 (0.69)	2.7 (0.64)	2.8 (0.60)		
T2	3.1 (0.73)	3.1 (0.74)	2.8 (0.74)		
Health management				8.67 (2, 69)	<.001
T1	2.4 (0.59)	2.2 (0.55)	2.3 (0.69)		
T2	2.9 (0.59)[a]	2.7 (0.60)[a]	2.3 (0.70)[b]		
Heart-healthy information, mean (SD)				2.68 (2, 69)	.08
T1	43.0 (4.04)	41.9 (4.04)	38.6 (6.49)		
T2	45.5 (3.03)	44.0 (3.01)	41.6 (4.12)		
Heart-healthy motivation, mean (SD)					
Overall				8.54 (2, 69)	<.001
T1	4.4 (0.80)	4.6 (0.74)	4.4 (0.56)		
T2	5.5 (0.60)[a]	5.2 (0.80)[a]	4.7 (0.76)[b]		
Effort or importance				9.45 (2, 69)	<.001
T1	3.8 (1.22)	4.3 (1.24)	3.9 (1.07)		
T2	5.5 (0.73)[a]	5.4 (0.93)[a]	4.5 (1.15)[b]		
Interest or enjoyment				10.48 (2, 69)	<.001
T1	3.7 (1.24)	4.1 (1.21)	4.0 (0.95)		
T2	5.5 (1.02)[a]	5.1 (1.09)[b]	4.4 (1.09)[c]		
Perceived choice				1.83 (2, 69)	.17
T1	4.7 (0.70)	4.6 (0.61)	4.5 (0.78)		
T2	5.0 (0.65)	4.9 (0.60)	4.6 (0.87)		
Perceived competence				3.28 (2, 69)	.04

	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	*F* test (*df*)	*P* value
T1	3.5 (1.28)	3.9 (1.13)	3.8 (0.83)		
T2	4.8 (1.04)[a]	4.7 (1.16)[a,b]	4.2 (1.13)[b]		
Value or usefulness				1.85 (2, 69)	.16
T1	5.9 (0.95)	6.0 (0.80)	5.7 (0.81)		
T2	6.5 (0.67)	6.1 (1.21)	5.9 (0.95)		
Heart-healthy self-efficacy for diet, mean (SD)				4.87 (2, 69)	.01
T1	71.5 (13.94)	71.2 (13.79)	75.4 (11.86)		
T2	79.8 (11.48)[a]	77.3 (10.62)[a]	73.1 (12.89)[b]		
Heart-healthy self-efficacy for exercise, mean (SD)				5.48 (2, 69)	.006
T1	37.4 (11.26)	38.9 (10.25)	36.2 (8.32)		
T2	40.3 (10.07)[a]	43.5 (7.43)[a]	34.8 (10.32)[b]		

[a]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[b]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[c]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

Regarding the subscales of heart-healthy behavior, significant group differences were observed in physical activity and weight control ($F_{2,69}$=4.02, *P*=.02), dietary habits ($F_{2,69}$=5.04, *P*=.009), and health management ($F_{2,69}$=8.67, *P*<.001; Table 2). Specifically, both the hybrid and mobile groups showed significantly greater increases in physical activity and weight control (hybrid vs control: MD 0.38, SD 0.17; *P*=.03; mobile vs control: MD 0.45, SD 0.17; *P*=.01), as well as health management (hybrid vs control: MD 0.62, SD 0.16; *P*<.001; mobile vs control: MD 0.49, SD 0.16; *P*=.002), compared with the control group. Moreover, the hybrid group (MD 0.42, SD 0.13; *P*=.002), but not the mobile group, demonstrated significantly greater improvements in dietary habit scores compared with the control group.

Regarding the subscales of heart-healthy motivation, significant group differences were observed in effort or importance ($F_{2,69}$=9.45, *P*<.001), interest or enjoyment ($F_{2,69}$=10.48, *P*<.001), and perceived competence ($F_{2,69}$=3.28, *P*=.04; Table 2). The hybrid group demonstrated significantly greater increases in interest or enjoyment than both the mobile (MD 0.54, SD 0.27; *P*=.045) and control groups (MD 1.21, SD 0.27; *P*<.001), while the mobile group also showed a significantly greater increase than the control group (MD 0.67, SD 0.26; *P*=.01). Both the hybrid (MD 1.00, SD 0.25; *P*<.001) and mobile (MD 0.79, SD 0.24; *P*=.002) groups showed significantly greater increases in effort or importance compared to the control group; however, there were no significant differences between the hybrid and mobile groups. Finally, the hybrid group (MD 0.70, SD 0.28; *P*=.01), but not the mobile group, demonstrated a significantly greater increase in perceived competence compared to the control group.

Cardiovascular Parameters

Table 3 summarizes the differential effects of the 3 groups on cardiovascular parameters over 12 weeks. Significant differences were observed between the 3 groups in fasting glucose levels ($F_{2,69}$=3.90, *P*=.03), with significant differences between the hybrid and control groups (*P*=.02) and the mobile and control groups (*P*=.02). However, there were no significant differences in the other parameters among the 3 groups.

Table 3. Effectiveness of a hybrid community-based heart-healthy lifestyle intervention on cardiovascular parameters (N=75). T1 and T2 refer to the baseline time point and the time point after 12 weeks, respectively.

	Hybrid group (n=25)	Mobile group (n=25)	Control group (n=25)	*F* test (*df*)	*P* value
BMI (kg/m^2), mean (SD)				0.82 (2, 69)	.45
T1	28.2 (4.84)	26.6 (3.73)	26.2 (4.14)		
T2	27.9 (4.79)	26.3 (3.93)	26.2 (4.01)		
WC[a] (cm), mean (SD)				0.20 (2, 69)	.82
T1	92.1 (10.47)	88.9 (10.39)	87.8 (11.26)		
T2	89.5 (9.39)	87.0 (10.47)	85.3 (10.73)		
Systolic BP[b], mean (SD)				0.07 (2, 69)	.93
T1	127.0 (18.36)	124.3 (19.10)	124.1 (18.77)		
T2	123.8 (14.77)	121.3 (17.15)	120.3 (12.30)		
Fasting glucose, mean (SD)				3.90 (2, 69)	.03
T1	93.9 (13.62)	92.9 (12.31)	92.2 (10.47)		
T2	90.6 (7.90)[c]	89.2 (8.98)[c]	93.2 (10.12)[d]		
LDL[e] cholesterol, mean (SD)				0.38 (2, 69)	.69
T1	128.0 (26.67)	123.5 (35.48)	126.0 (34.77)		
T2	129.9 (21.88)	122.5 (30.42)	123.0 (36.08)		
HDL[f] cholesterol, mean (SD)				0.78 (2, 69)	.46
T1	53.2 (14.81)	53.6 (13.20)	55.4 (14.86)		
T2	55.0 (13.83)	57.7 (13.70)	56.9 (13.18)		
Triglycerides, mean (SD)				0.57 (2, 69)	.57
T1	248.3 (372.18)	169.8 (86.41)	155.7 (104.73)		
T2	163.1 (83.39)	130.5 (65.80)	145.0 (85.40)		

[a]WC: waist circumference.

[b]BP: blood pressure.

[c]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[d]Different superscripts indicate a statistically significant difference by Tukey least significant difference multiple comparison.

[e]LDL: low-density lipoprotein.

[f]HDL: high-density lipoprotein.

Discussion

Principal Results

We found that, compared to the control group, both the hybrid and mobile intervention groups demonstrated significantly greater improvements in heart-healthy behavioral outcomes, including composite heart-healthy behavior, its theoretical predictors (heart-healthy motivation and self-efficacy), and fasting glucose levels following the 12-week hybrid intervention for individuals at cardiovascular risk. In particular, the hybrid group—unlike the mobile group—showed significantly greater improvement in dietary behavior, a subscale of composite heart-healthy behavior, compared with the control group, and also demonstrated significantly greater improvements in interest or enjoyment, a core subscale of intrinsic motivation, than the mobile and control groups [45]. These findings suggest that the hybrid community-based heart-healthy lifestyle intervention—integrating the mobile app and motivational interviewing—demonstrated overall effectiveness comparable to the mobile app alone, while yielding greater improvements in dietary behavior and core intrinsic motivation.

The hybrid and mobile groups were more likely to increase the composite score of heart-healthy behaviors than the control group after the 12-week intervention, with no significant difference observed between the 2 groups. However, the hybrid group—unlike the mobile group—showed significantly greater improvement in dietary behavior, a subscale of composite heart-healthy behavior, compared with the control group. The finding regarding composite heart-healthy behavior suggests that motivational interviewing may not have demonstrated superiority over the use of a mobile app alone, indicating that the “My HeartHELP” mobile app itself may effectively facilitate overall improvements in heart-healthy behaviors. The “My HeartHELP” app uniquely incorporates 6 key heart-healthy

behaviors simultaneously, which may significantly affect the composite score of heart-healthy behaviors. Empirically, this positive finding may be explained by the fact that the "My HeartHELP" app incorporates evidence-based behavioral change strategies [10,13,14,23], including text messaging for enhancing information, encouraging self-monitoring, and providing feedback or reinforcement in line with individuals' behavioral goal setting.

To date, few mHealth studies on cardiovascular health have primarily relied on theory-based smartphone apps targeting multiple heart-healthy behaviors, as in our study. Most studies, by contrast, have used a variety of mHealth modalities, such as commercial apps, telephone, web-based, email, and SMS text messaging, each corresponding to a single specific behavior, such as physical activity, weight loss, or smoking [20]. Furthermore, most previous studies have predominantly emphasized health outcomes as direct indicators of effectiveness [6], without accounting for the internal validity of mHealth interventions through behavioral change mechanisms. In this regard, our study may be unique in its integration of a mobile app with evidence-based behavioral change strategies specifically designed to promote multiple heart-healthy behaviors and in its comprehensive evaluation of both behavioral changes and health outcomes, including cardiovascular parameters [14].

Meanwhile, the hybrid group, but not the mobile group, showed a significantly greater improvement in the subscale for dietary behavior after 12 weeks, compared to the control group. This finding indicates that motivational interviewing may have had an additional effect on changing dietary habits through the use of the "My HeartHELP" app. Motivational interviewing has traditionally demonstrated significant effectiveness in promoting behavioral changes in substance use, especially in diverse psychological contexts [17]. Dietary behavior involves adherence to diverse dietary guidelines and the consumption of various food groups [46], making effective intervention challenging without considering the psychological and social contexts of individuals [47]. In this regard, motivational interviewing was incorporated in the present study to mitigate the limitations of mobile apps in facilitating close expert interaction. This approach may be effective in facilitating a person-centered understanding of the psychosocial context of complex eating behavior and allows for personalized adjustments to eating behaviors, thus enhancing the intervention's effectiveness. This highlights motivational interviewing's particular efficacy for complex behavioral domains that require nuanced, individualized support beyond what a mobile app alone can provide.

Our findings demonstrated that both the hybrid and mobile groups exhibited greater increases in overall scores of the heart-healthy motivation compared to the control group, with no significant difference between the hybrid and mobile groups. This current study did not fully support the hypothesis that motivational interviewing would be more effective than using a mobile app alone [48,49]. Nevertheless, the hybrid group demonstrated significantly greater increases in a core subscale of intrinsic motivation—interest or enjoyment [50]—than the mobile group. In this context, the above findings may not entirely rule out the possibility that motivational interviewing could serve as an effective intervention component for strengthening intrinsic motivation [24,51], potentially offering advantages beyond the use of a mobile app alone. This finding also aligns with the conceptualization of intrinsic motivation as measured by the 4 items for interest or enjoyment in the Situational Motivation Scale [45].

However, the lack of overall statistical superiority of motivational interviewing may be explained from 2 perspectives. First, the "My HeartHELP" app incorporated substantial motivational elements, including self-monitoring, automated and tailored feedback, and goal attainment encouragement [52]. Second, the limited dosage of the motivational interviewing—4 sessions over 12 weeks—may have been insufficient to produce statistically significant additive effects across all motivational outcomes [53]. In line with this dosage, the quality and intensity of motivational interviewing delivered in this study may have constrained its effectiveness, beyond that of the mobile app alone, in fostering individuals' autonomy to adjust behavioral goal settings, as both the mobile and hybrid groups adhered to the same predefined behavioral goals. Future research should optimize the dosage of motivational interviewing and enhance its quality of autonomy-supportive interventions, thereby more strategically and intensively enabling participants to independently design their activities and personalize goal-setting, as emphasized in self-determination theory [54].

Theoretically grounded in the IMB model [23], the observed increase in heart-healthy motivation in both the hybrid and mobile groups can be interpreted as evidence that enhanced motivation may have served as a foundational mechanism strengthening the mediating variable—heart-healthy self-efficacy—which, in turn, contributed to improvements in heart-healthy behavior (Multimedia Appendix 1). Consistent with this theoretical pathway, our path analytic data (not shown) indicated that increases in overall heart-healthy motivation scores were significantly and positively associated with increases in self-efficacy for diet (β=.32, *P*=.02) and exercise (β=.38, *P*=.005), as well as with improvements in heart-healthy behaviors (β=.27, *P*=.04). These findings may be attributable to the intervention strategies used in this study, which fostered self-directed, health-oriented internalization through self-monitoring and cognitive reappraisal—encompassing the daily input and reflection on 6 core heart-healthy behaviors via the app and motivational interviewing that encouraged participants to recognize progress and use self-management. Subsequently, self-efficacy for diet and exercise may have been enhanced through text messaging and tailored feedback reflecting individual success (eg, mastery experience) delivered via "My HeartHELP" [14], along with professional encouragement and persuasive support provided through motivational interviewing. Collectively, these cognitive and behavioral mechanisms may have contributed to the observed improvements in overall heart-healthy behaviors in either the hybrid or mobile group.

In our findings, the significantly increased self-efficacy for diet and exercise after 12 weeks of using the "My HeartHELP" app may be explained by the efficacy of the behavioral strategies embedded in the app. Goal-setting and self-monitoring can

contribute to mastery experiences [55] by allowing participants to observe their own progress and achieve incremental success, reinforcing their belief in their ability to sustain heart-healthy behaviors [55]. Moreover, the "My HeartHELP" app's personalized feedback messaging delivered daily, weekly, and monthly based on individuals' behavioral outcomes could function as a form of verbal persuasion, as also described by Bandura [55], reinforcing self-efficacy through experts' positive reinforcement and targeted feedback on behavioral progress. Moreover, the "My HeartHELP" app delivered 43 text messages on behavioral skills, informing users on how to apply these skills in daily life. Mobile apps are currently limited in their ability to sensitively recognize and address individuals' emotional states. However, the 4 core principles of motivational interviewing (ie, expressing empathy, developing discrepancies, rolling with resistance, and supporting self-efficacy) are thought to be effective in addressing such emotional states [17]. Previous studies have suggested that motivational interviewing, as a behavioral intervention, can be highly effective in enhancing individuals' self-efficacy [51]. Nevertheless, the lack of differences observed between the hybrid and mobile-only groups may be due to the strong self-efficacy effects of the mobile app overshadowing the effects of motivational interviewing.

We found no significant differences in heart-healthy information levels among the 3 groups following the 12-week intervention. All participants were provided with educational materials on cardiovascular health; the mobile group received cardiovascular information via text messaging, while the hybrid group received both text messages and individualized information. However, an analysis of the item-response rate revealed that participants in all 3 groups continued to provide incorrect responses to certain items on the Heart-Healthy Information Questionnaire at posttest—items that were nearly identical to those answered incorrectly at baseline [34]—despite completing the 12-week intervention. These findings suggest that interventionists should proactively address specific questionnaire items that participants struggled with at the baseline assessment. Integrating targeted educational reinforcement into the intervention design may be necessary to enhance knowledge acquisition and retention.

Strengths and Limitations of This Study

This study underscored the potential of mHealth within a community setting for behavioral lifestyle interventions by optimizing user-interventionist interactions through tailored feedback delivered via in-app text messaging. To the best of our knowledge, this study is the first to reveal the effectiveness of a hybrid approach combining mobile apps and motivational interviewing counseling on heart-healthy behavioral changes, compared to mobile apps alone, within a community setting. Moreover, this indicated that the mHealth intervention using the mobile app effectively facilitated significant changes in heart-healthy behaviors, and the validity of these changes was confirmed through significant improvements in behavioral predictors based on the IMB model. This is particularly meaningful, as it establishes the internal validity of the intervention, demonstrating that the observed behavioral changes were systematically driven by theoretical constructs.

Nevertheless, this study has several limitations. First, dropouts (2 subjects in the hybrid group and 1 in the mobile group) may have led to a bias influencing the validity of the results, even though we performed an intention-to-treat analysis. Second, the absence of participant blinding in this study's design may have introduced bias into this study's findings. Third, because all the participants in this study were Korean, the results cannot be generalized to other ethnic groups. Fourth, because the survey was conducted online, participants with lower digital literacy were likely underrepresented. This is reflected in the socioeconomic profile of our participants, with a high proportion of college-educated individuals (64/75, 85.3%) compared with the national average in South Korea [36]. These characteristics may have facilitated greater engagement with the mHealth intervention and should be considered when generalizing the findings to populations with lower education levels or limited digital access. Fifth, this study lacks a qualitative component, which limits the understanding of why the intervention was effective from the participants' perspective. Qualitative feedback, such as through user interviews, could have provided invaluable insight into which app features were most engaging or how the motivational interviewing sessions were perceived. Sixth, given that this study targeted multiple healthy lifestyle behaviors, appropriate behavioral and self-efficacy measures specifically designed for the comprehensive assessment of heart-healthy behaviors are lacking. Moreover, the measure of heart-healthy motivation used in this study excluded 2 subscales—relatedness and pressure or tension—because their items were deemed inappropriate for the context of adopting heart-healthy behaviors, although these dimensions, particularly relatedness, may play an essential role. The rationale for this exclusion has been described above in the Methods section. The instrument was developed and refined through a rigorous translation and back-translation process; however, the absence of a formal validation process, including factor analysis and psychometric testing, warrants cautious interpretation of the present findings. Therefore, a subsequent validation study should be conducted among the Korean population. In light of these limitations, future research should prioritize the development and validation of robust assessment tools that accurately capture the multidimensional nature of heart-healthy behavioral constructs, thereby ensuring a more precise evaluation of intervention outcomes.

Implications for Policy and Practice

The use of a mobile app alone may be effective in facilitating changes in heart-healthy behaviors when integrated with evidence-based behavioral strategies. Given the limitations of expert-centered cardiovascular care in community settings, a mobile heart-healthy intervention led by nurses or other health professionals may serve as an efficient and accessible alternative. Although motivational interviewing did not demonstrate universal superiority over the mobile app alone, it remains a valuable component for behavioral interventions. More importantly, the hybrid approach used in this study demonstrated the potential of motivational interviewing not only to enhance intervention effects—particularly in improving intricate heart-healthy dietary behaviors—but also to promote sustainability by reinforcing intrinsic motivation, specifically

interest or enjoyment [45]. We therefore strongly recommend integrating motivational interviewing into mHealth interventions, especially when targeting subtle behavioral changes and fostering core intrinsic motivation, as supported by our findings on the motivation subscales. Furthermore, the intervention could be further strengthened by integrating periodic counseling sessions using the motivational interviewing modality alongside the mobile intervention. Therefore, workforce training programs should be implemented to equip health professionals with the skills necessary to deliver mHealth interventions and motivational interviewing as components of cardiovascular health promotion. This approach could enhance the effectiveness and accessibility of heart health interventions in community settings.

Recommendations for Further Research

Future studies should explore the long-term effects of mobile apps and their combinations with motivational interviewing on heart-healthy behavioral outcomes. Additionally, optimizing the frequency and delivery of motivational interviewing sessions should be considered, particularly within the constraints of limited resources in community settings, to enhance the sustainability and effectiveness of behavioral interventions.

Conclusions

The hybrid community-based heart-healthy lifestyle intervention—integrating the mobile app and motivational interviewing—demonstrated comparable overall effectiveness to the mobile app alone, yet achieved greater improvements in intrinsic motivation (interest or enjoyment) and dietary behavior. These findings highlight the potential of mHealth apps as practical, stand-alone tools to promote cardiovascular health, particularly in community settings with limited access to in-person professional support. However, incorporating motivational interviewing may further enhance internalized motivation and sustain complex behavior changes over time. Health professionals can therefore adopt mHealth either independently or in combination with motivational interviewing to optimize heart-healthy behavioral outcomes. Future studies should optimize integration strategies to enhance effectiveness and evaluate the long-term sustainability of such hybrid approaches.

Funding

This research was supported by the National Research Foundation of Korea (NRF) grant from the Korean Government (No. NRF-2019R1A2C1004116) and that from the Ministry of Science and ICT (Information and Communication Technology; No. RS-2024-00336847).

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Conceptualization: JC
Data curation: YS, SN
Formal analysis: YS
Funding acquisition: JC
Methodology: JC
Project administration: JC, SN
Resources: YS, SN
Supervision: JC, JL
Validation: JC, JL
Visualization: YS
Writing – original draft: JC
Writing – review & editing: JC, YS, SN, JL

Conflicts of Interest

None declared.

Multimedia Appendix 1
The hypothetical model based on constructs of the IMB model. IMB: Information-Motivation-Behavior. Reprinted with permission from the American Psychological Association. 2025 American Psychological Association. License obtained.
[PNG File , 53 KB - mhealth_v14i1e76521_app1.png]

Multimedia Appendix 2
CONSORT-EHEALTH (V1.6.1) checklist.
[PDF File (Adobe PDF File), 1318 KB - mhealth_v14i1e76521_app2.pdf]

Multimedia Appendix 3
Expanded outline of the 12-week hybrid community-based heart-healthy lifestyle intervention.
[PNG File , 1008 KB - mhealth_v14i1e76521_app3.png]

References

1. World Heart Report 2023: confronting the world's number one killer. World Heart Federation. Geneva, Switzerland: World Heart Federation; 2023. URL: https://world-heart-federation.org/wp-content/uploads/World-Heart-Report-2023.pdf [accessed 2025-12-20]
2. Cardiovascular diseases (CVDs). World Health Organization. Geneva, Switzerland: World Health Organization; 2025. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2025-12-20]
3. Wu J, Feng Y, Zhao Y, Guo Z, Liu R, Zeng X, et al. Lifestyle behaviors and risk of cardiovascular disease and prognosis among individuals with cardiovascular disease: a systematic review and meta-analysis of 71 prospective cohort studies. Int J Behav Nutr Phys Act 2024;21(1):42 [FREE Full text] [doi: 10.1186/s12966-024-01586-7] [Medline: 38650004]
4. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. Prev Med 2012;55(3):163-170. [doi: 10.1016/j.ypmed.2012.06.017] [Medline: 22735042]
5. Fang J, Moore L, Loustalot F, Yang Q, Ayala C. Reporting of adherence to healthy lifestyle behaviors among hypertensive adults in the 50 states and the District of Columbia, 2013. J Am Soc Hypertens 2016;10(3):252-262.e3 [FREE Full text] [doi: 10.1016/j.jash.2016.01.008] [Medline: 26851000]
6. Sequi-Dominguez I, Alvarez-Bueno C, Martinez-Vizcaino V, Fernandez-Rodriguez R, Del Saz Lara A, Cavero-Redondo I. Effectiveness of mobile health interventions promoting physical activity and lifestyle interventions to reduce cardiovascular risk among individuals with metabolic syndrome: systematic review and meta-analysis. J Med Internet Res 2020;22(8):e17790 [FREE Full text] [doi: 10.2196/17790] [Medline: 32865503]
7. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile Health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
8. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR mHealth uHealth 2020;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
9. Khunti K, Griffin S, Brennan A, Dallosso H, Davies MJ, Eborall HC, et al. Promoting physical activity in a multi-ethnic population at high risk of diabetes: the 48-month PROPELS randomised controlled trial. BMC Med 2021;19(1):130 [FREE Full text] [doi: 10.1186/s12916-021-01997-4] [Medline: 34078362]
10. Burke LE, Sereika SM, Parmanto B, Bizhanova Z, Kariuki JK, Cheng J, et al. Effect of tailored, daily feedback with lifestyle self-monitoring on weight loss: the SMARTER randomized clinical trial. Obesity (Silver Spring) 2022;30(1):75-84. [doi: 10.1002/oby.23321] [Medline: 34898011]
11. Giebel GD, Speckemeier C, Abels C, Plescher F, Börchers K, Wasem J, et al. Problems and barriers related to the use of digital health applications: scoping review. J Med Internet Res 2023;25:e43808 [FREE Full text] [doi: 10.2196/43808] [Medline: 37171838]
12. Müssener U. Digital encounters: human interactions in mHealth behavior change interventions. Digit Health 2021;7:20552076211029776 [FREE Full text] [doi: 10.1177/20552076211029776] [Medline: 34262783]
13. Artinian NT, Fletcher GF, Mozaffarian D, Kris-Etherton P, Van Horn L, Lichtenstein AH, American Heart Association Prevention Committee of the Council on Cardiovascular Nursing. Interventions to promote physical activity and dietary lifestyle changes for cardiovascular risk factor reduction in adults: a scientific statement from the American Heart Association. Circulation 2010;122(4):406-441 [FREE Full text] [doi: 10.1161/CIR.0b013e3181e8edf1] [Medline: 20625115]
14. Choo J, Noh S, Shin Y. Evaluating feasibility and acceptability of the "My HeartHELP" mobile app for promoting heart-healthy lifestyle behaviors: mixed methods study. JMIR Form Res 2025;9:e66108 [FREE Full text] [doi: 10.2196/66108] [Medline: 40315608]
15. Patnode CD, Redmond N, Iacocca MO, Henninger M. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults without known cardiovascular disease risk factors: updated evidence report and systematic review for the US preventive services task force. JAMA 2022;328(4):375-388. [doi: 10.1001/jama.2022.7408] [Medline: 35881116]
16. Polcin DL, Sterling J, Brown T, Brown M, Buscemi R, Korcha R. Client and therapist views about intensive and standard motivational interviewing. J Contemp Psychother 2015;45(3):167-176 [FREE Full text] [doi: 10.1007/s10879-014-9280-1] [Medline: 26185335]
17. Miller W, Rollnick S. Motivational Interviewing: Helping People Change. 3rd ed. New York (NY): Guilford Press; 2012.
18. Frost H, Campbell P, Maxwell M, O'Carroll RE, Dombrowski SU, Williams B, et al. Effectiveness of motivational interviewing on adult behaviour change in health and social care settings: a systematic review of reviews. PLoS One 2018;13(10):e0204890 [FREE Full text] [doi: 10.1371/journal.pone.0204890] [Medline: 30335780]
19. Rotger JM, Cabré V. Therapeutic alliance in online and face-to-face psychological treatment: comparative study. JMIR Ment Health 2022;9(5):e36775 [FREE Full text] [doi: 10.2196/36775] [Medline: 35499910]

20. Cajita MI, Zheng Y, Kariuki JK, Vuckovic KM, Burke LE. mHealth technology and CVD risk reduction. Curr Atheroscler Rep 2021;23(7):36. [doi: 10.1007/s11883-021-00927-2] [Medline: 33983491]
21. Yun YH, Kang E, Cho YM, Park SM, Kim Y, Lee H, et al. Efficacy of an electronic health management program for patients with cardiovascular risk: randomized controlled trial. J Med Internet Res 2020;22(1):e15057 [FREE Full text] [doi: 10.2196/15057] [Medline: 32012053]
22. Ogedegbe G, Teresi JA, Williams SK, Ogunlade A, Izeogu C, Eimicke JP, et al. Home blood pressure telemonitoring and nurse case management in Black and Hispanic patients with stroke: a randomized clinical trial. JAMA 2024;332(1):41-50. [doi: 10.1001/jama.2024.6609] [Medline: 38842799]
23. Fisher W, Fisher J, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Social Psychological Foundations of Health and Illness. New York (NY): Wiley; 2003:82-106.
24. Nuss K, Moore K, Marchant T, Courtney JB, Edwards K, Sharp JL, et al. The combined effect of motivational interviewing and wearable fitness trackers on motivation and physical activity in inactive adults: a randomized controlled trial. J Sports Sci 2023;41(1):45-55. [doi: 10.1080/02640414.2023.2195228] [Medline: 36966352]
25. Burke LE, Choo J, Music E, Warziski M, Styn MA, Kim Y, et al. PREFER study: a randomized clinical trial testing treatment preference and two dietary options in behavioral weight management — rationale, design and baseline characteristics. Contemp Clin Trials 2006;27(1):34-48. [doi: 10.1016/j.cct.2005.08.002] [Medline: 16233990]
26. Hansson L, Hedner T, Dahlöf B. Prospective randomized open blinded end-point (PROBE) study. A novel design for intervention trials. Blood Press 1992;1(2):113-119 [FREE Full text] [doi: 10.3109/08037059209077502] [Medline: 1366259]
27. Eysenbach G. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
28. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation,Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. Circulation 2002;106(25):3143-3421. [Medline: 12485966]
29. Park S. Effects of a Community-Based Nursing Program for Promoting Cardiometabolic Health Behaviors Among Postmenopausal Middle-Aged Women: on a Basis of the Extended Theory of Planned Behavior [dissertation]. Seoul (KR): Korea University; 2019.
30. Abrams KR, Gillies CL, Lambert PC. Meta-analysis of heterogeneously reported trials assessing change from baseline. Stat Med 2005;24(24):3823-3844. [doi: 10.1002/sim.2423] [Medline: 16320285]
31. Symer MM, Abelson JS, Milsom J, McClure B, Yeo HL. A mobile health application to track patients after gastrointestinal surgery: results from a pilot study. J Gastrointest Surg 2017;21(9):1500-1505. [doi: 10.1007/s11605-017-3482-2] [Medline: 28685388]
32. Miller W, Rollnick S, Moyers T. Motivational interviewing. 2021. URL: https://psychwire.com/motivational-interviewing [accessed 2025-12-20]
33. Kang S. The validity and reliability of a lifestyle evaluation tool for patients with metabolic syndrome. J Korean Acad Fundam Nurs 2010;17(4):487-497 [FREE Full text]
34. Choo J, Noh S, Moon J, Shin Y. Development and psychometric testing of the Heart-Healthy Information Questionnaire. Eur J Cardiovasc Nurs 2023;22(3):299-310. [doi: 10.1093/eurjcn/zvac055] [Medline: 35766172]
35. McAuley E, Duncan T, Tammen VV. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: a confirmatory factor analysis. Res Q Exerc Sport 1989;60(1):48-58. [doi: 10.1080/02701367.1989.10607413] [Medline: 2489825]
36. Korea Disease Control and Prevention Agency. The eighth korea national health and nutrition examination survey (KNHANES VIII) Internet. Cheongju (KR): KDCA; 2021. URL: https://knhanes.kdca.go.kr/knhanes/eng/main.do [accessed 2026-01-03]
37. Cocca A, Veulliet N, Niedermeier M, Drenowatz C, Cocca M, Greier K, et al. Psychometric parameters of the intrinsic motivation inventory adapted to physical education in a sample of active adults from Austria. Sustainability 2022;14(20):13681. [doi: 10.3390/su142013681]
38. Gibbens B. Measuring student motivation in an introductory biology class. Am Biol Teach 2019;81(1):20-26. [doi: 10.1525/abt.2019.81.1.20]
39. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]
40. Shin C, Lach H. Nutritional issues of Korean Americans. Clin Nurs Res 2011;20(2):162-180. [doi: 10.1177/1054773810393334] [Medline: 21160079]
41. Choo J, Lee J, Cho J, Burke LE, Sekikawa A, Jae SY. Effects of weight management by exercise modes on markers of subclinical atherosclerosis and cardiometabolic profile among women with abdominal obesity: a randomized controlled trial. BMC Cardiovasc Disord 2014;14:82 [FREE Full text] [doi: 10.1186/1471-2261-14-82] [Medline: 25011384]
42. Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Jager KJ. Intention to treat and per protocol analysis in clinical trials. Nephrology (Carlton) 2020 Jul;25(7):513-517. [doi: 10.1111/nep.13709] [Medline: 32147926]

43. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
44. Ministry of Health and Welfare. The Median Income Level of Korea. Sejong (KR): Ministry of Health and Welfare; 2024. URL: https://www.mohw.go.kr/menu.es?mid=a10708010900 [accessed 2026-01-03]
45. Guay F, Vallerand RJ, Blanchard C. On the assessment of situational intrinsic and extrinsic motivation: the Situational Motivation Scale (SIMS). Motiv Emotion 2000;24(3):175-213. [doi: 10.1023/a:1005614228250]
46. Nocella G, Srinivasan C. Adherence to WHO's nutrition recommendations in the UK: dietary patterns and policy implications from a national survey. Food Policy 2019;86:101719. [doi: 10.1016/j.foodpol.2019.05.002]
47. Stevenson RJ. Psychological correlates of habitual diet in healthy adults. Psychol Bull 2017;143(1):53-90. [doi: 10.1037/bul0000065] [Medline: 27618545]
48. Young HM, Miyamoto S, Dharmar M, Tang-Feldman Y. Nurse coaching and mobile health compared with usual care to improve diabetes self-efficacy for persons with type 2 diabetes: randomized controlled trial. JMIR mHealth uHealth 2020;8(3):e16665 [FREE Full text] [doi: 10.2196/16665] [Medline: 32130184]
49. Wong AKC, Bayuo J, Wong FKY, Chow KKS, Wong SM, Lau ACK. The synergistic effect of nurse proactive phone calls with an mHealth app program on sustaining app usage: 3-arm randomized controlled trial. J Med Internet Res 2023;25:e43678 [FREE Full text] [doi: 10.2196/43678] [Medline: 37126378]
50. Intrinsic motivation inventory: constructing the IMI for your study internet. Center for Self-Determination Theory. Rochester (NY): University of Rochester; 2022. URL: https://selfdeterminationtheory.org/wp-content/uploads/2022/02/IMI_Complete.pdf [accessed 2025-12-20]
51. Khadoura K, Shakibazadeh E, Mansournia M, Aljeesh Y, Fotouhi A. Effectiveness of motivational interviewing on medication adherence among Palestinian hypertensive patients: a clustered randomized controlled trial. Eur J Cardiovasc Nurs 2021;20(5):411-420. [doi: 10.1093/eurjcn/zvaa015] [Medline: 34009313]
52. Ambeba E, Ye L, Sereika S, Styn M, Acharya S, Sevick M, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43 [FREE Full text] [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
53. Lee WW, Choi K, Yum RW, Yu DS, Chair S. Effectiveness of motivational interviewing on lifestyle modification and health outcomes of clients at risk or diagnosed with cardiovascular diseases: a systematic review. Int J Nurs Stud 2016;53:331-341. [doi: 10.1016/j.ijnurstu.2015.09.010] [Medline: 26493130]
54. Self-determination theory. Wikipedia. 2025. URL: https://en.wikipedia.org/wiki/Self-determination_theory [accessed 2025-12-20]
55. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977;84(2):191-215. [doi: 10.1037//0033-295x.84.2.191] [Medline: 847061]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
IMB: Information-Motivation-Behavioral Skills
MD: mean difference
mHealth: mobile health

Edited by L Buis; submitted 25.Apr.2025; peer-reviewed by P Pichayapinyo, H Sadeghsalehi; comments to author 24.Sep.2025; revised version received 16.Oct.2025; accepted 17.Dec.2025; published 15.Jan.2026.

Please cite as:
Choo J, Shin Y, Noh S, Lee J
Effectiveness of a Hybrid Community-Based Heart-Healthy Lifestyle Intervention: Three-Arm Randomized Controlled Trial Integrating mHealth and Motivational Interviewing
JMIR Mhealth Uhealth 2026;14:e76521
URL: https://mhealth.jmir.org/2026/1/e76521
doi:10.2196/76521
PMID:41407528

Original Paper

A Brief Web-Based and Mobile Intervention of Intermittent Fasting With Meal Support for Weight Loss Among Adults With Overweight and Obesity in Japan: Pilot Randomized Controlled Trial

Takashi Noda[1], MPH; Tomonari Shimamoto[1], DPH; Kosuke Kiyohara[2], DPH; Mie Imanaka[3], MPH, RD; Norihiro Nishioka[1], MD, PhD; Yukiko Tateyama[1], PhD; Taku Iwami[1], MD, PhD

[1]Department of Preventive Services, School of Public Health, Graduate School of Medicine, Kyoto University, Kyoto, Japan

[2]Department of Food Science, Faculty of Home Economics, Otsuma Women's University, Tokyo, Japan

[3]Department of Health and Nutrition, Faculty of Nursing and Nutrition, The University of Shimane, Izumo, Japan

Corresponding Author:
Taku Iwami, MD, PhD
Department of Preventive Services
School of Public Health, Graduate School of Medicine
Kyoto University
Yoshida Konoe-cho
Sakyo-ku
Kyoto, 606-8315
Japan
Phone: 81 757534400
Email: iwami.taku.8w@kyoto-u.ac.jp

Abstract

Background: Intermittent fasting emerges as a promising dietary approach against obesity, offering a cost-effective strategy for implementation via web-based platforms. We developed a Brief Online Intermittent Fasting Program (OIF), featuring a self-administered, weekly 1-day fasting regimen with replacement meals delivery, online guidance, and app messaging to support adherence.

Objective: This pilot study aimed to assess the preliminary effectiveness, feasibility, and safety of the OIF on weight loss in adults with overweight and obesity in Japan. Secondary objectives were to assess its effects on body composition and metabolic markers.

Methods: This 12-week, 1:1 randomized controlled trial recruited adults with overweight and obesity (BMI from 23 to <35) in 1 university, 1 hospital, and 2 company offices. Participants were randomized into 2 groups stratified by sex and age (<40 or ≥40 years). The intervention group received very low-calorie (407 kcal) meal replacements for weekly intermittent fasting, online guidance via Zoom (Zoom Video Communications, Inc), and app messages encouraging fasting and healthy lifestyles. The control group received app messages promoting healthy lifestyles only. Interventions were administered by a nonblinded researcher. The primary outcome was the change in body weight after 12 weeks, analyzed using intention-to-treat principles and adjusted for sex, age, and baseline weight. Secondary outcomes encompassed body composition, blood pressure, biomarkers (eg, hemoglobin A_{1c}, triglycerides, and cholesterol), quality of life, physical activity, intervention adherence, and adverse events.

Results: A total of 57 individuals were enrolled (28 in the intervention group and 29 in the control group). At 12 weeks, 25 participants in the intervention group and 27 participants in the control group completed follow-up. The baseline median weight was 75.8 (IQR 68.3-80.6) kg for the intervention group and 74.8 (IQR 69.8-81.8) kg for the control group. The mean weight change was –0.9 (SD 1.9) kg in the intervention group and +0.6 (SD 1.4) kg in the control group. The adjusted between-group difference in weight change was statistically significant at –1.6 (95% CI –2.5 to –0.8) kg. Fat mass change was not statistically significant (–0.1, 95% CI –1.3 to 1.4 kg), but muscle mass reduction was implied (–1.3, 95% CI –2.5 to –0.2 kg). Intervention adherence was 79% (22/28) in the intervention group. No serious adverse events were reported, and there were no significant changes in key biomarkers, such as hemoglobin A_{1c} or quality of life.

Conclusions: The OIF demonstrated effectiveness in promoting modest weight loss among adults with overweight and obesity over 12 weeks, with high feasibility and safety indicated by low dropout rates and absence of serious adverse events. However,

the observed reduction in muscle mass indicates a need for program refinement, such as incorporating exercise guidance, to optimize health outcomes.

Trial Registration: UMIN-CTR UMIN000050437; https://tinyurl.com/4x5h2t2x

(JMIR Mhealth Uhealth 2026;14:e58930) doi:10.2196/58930

KEYWORDS

intermittent fasting; eHealth; mHealth; mobile health; weight loss; diet; overweight; obesity; randomized controlled trial

Introduction

Obesity is a well-known public health challenge, which increases the risk of cardiovascular diseases and mortality [1]. According to the World Health Organization, at least 2.8 million people die each year as a result of being overweight or obese [2], with the obesity epidemic still escalating at an alarming rate [3]. Moreover, the recent COVID-19 pandemic has further deteriorated our lifestyles and made the development of remote obesity interventions, using the internet, apps, and other web-based technologies, an even more urgent issue [4-6]. Although many web-based programs have been developed, with some degree of effectiveness acknowledged [7,8], few have been successfully implemented in society to yield significant results. Also, criticism exists toward the previous clinical trial approaches, which have focused too much on internal validity [9], indicating a demand for the development of highly feasible online weight loss programs that are easily applied and disseminated into society.

Recently, intermittent fasting (IF) has emerged as a popular strategy for combating obesity [10]. IF is a simple dietary restriction regimen that incorporates periodic fasting (energy restriction of less than 600 kcal) into normal eating habits [11,12]. Recent systematic reviews and meta-analyses have demonstrated the effectiveness of IF for weight loss compared to ad libitum eating [13]. The mechanisms behind weight loss with IF are related to energy restriction, increased fat metabolism, and improved glucose metabolism [14], which are common to any dietary restriction. IF may also provide cardiovascular health benefits. For example, studies with women with obesity show that a weekly IF intervention over 2 months significantly reduced resting heart rate, blood glucose, insulin, and homocysteine levels. These effects are likely associated with reductions in adipokines, including leptin, IL-6, tumor necrosis factor-alpha, and insulin-like growth factor-1, suggesting that IF could mitigate cardiovascular disease risk [15,16]. Additional benefits reported in recent studies include decreased hunger, reduced inflammation, improved cellular repair, regulated hormone production, enhanced immune function, adjusted circadian rhythms, and a more diverse gut microbiome [14,17-21]. Therefore, IF holds significant potential as a promising nonpharmacological approach to improving health at the population level [18]. We particularly focused on the simplicity and feasibility of IF, because a truly successful weight loss approach needs to be simple and sustainable, capable of being incorporated into daily life as lifelong dietary customs [18,21]. Since IF does not require detailed examination of food content or energy calculations, it is easy to implement with brief web-based guidance, having a high possibility of being applied as a cost-effective population approach. However, to the best of our knowledge, no studies have implemented IF as a web-based intervention, leaving a potentially significant solution to obesity unexplored.

To assess the effect of web-based IF administration, we have developed the “Brief Online Intermittent Fasting Program (OIF),” which features self-led fasting practice supported with meal delivery, online guidance, and app messages. The purposes of this pilot study were to evaluate the preliminary effectiveness, feasibility, and safety of the 12-week OIF on weight loss in adults with overweight and obesity in Japan. Secondary objectives were to assess its effects on body composition and metabolic markers.

Methods

Study Design

This pilot study is a 1:1 parallel-group randomized controlled trial (RCT) comparing the intervention group following the OIF with the control group receiving minimal care.

Participants and Setting

Eligibility Criteria

In order to benefit from weight reduction and ensure the safe and stable administration of the OIF, the following eligibility criteria were established.

The inclusion criteria are as follows: individuals with overweight or obesity and with a BMI between 23 and under 35, adults aged 20 years to less than 65 years at the time of consent, those capable of communicating in Japanese, owners of a smartphone with iOS or Android, those who were familiar with using apps on a smartphone and using email and Zoom (Zoom Video Communications, Inc) on either a smartphone or a personal computer, those willing to visit one of the research facilities twice, and consenting to provide blood samples through self-puncture at the measurements. We defined overweight as having a BMI of 23 or above, while the global standard for overweight is a BMI of 25 or above. This adjustment aligns with guidelines from the National Institute for Health and Care Excellence, which recommend lowering the overweight threshold to 23 for Asian populations to enhance metabolic disease prevention in populations with lower obesity prevalence [22].

The exclusion criteria include any of the following: individuals with a history of heart disease, kidney disease, psychiatric disorders, or any serious medical conditions, individuals using diabetes medication (to minimize the risk of hypoglycemia), smokers or heavy drinkers (consuming alcohol on average more

than 4 drinks per day or 14 drinks per week), night shift workers, individuals who may face significant environmental changes during the trial period, individuals with eating disorders, food allergies or alcohol dependence who require specific dietary supervision, women who are breastfeeding, pregnant, or planning to become pregnant during the study, lean individuals with a body fat percentage below 10% for men and 20% for women, athletes engaging in more than 12 hours of training or sports practice per week or those belonging to university sports clubs or professional teams aiming for competition, and any other individuals deemed unsuitable for participation by the researchers.

Setting and Recruitment Procedure

Participant recruitment, measurement, and registration were conducted at Kyoto University and 3 collaborating facilities in Japan: Kusaka Hospital, Cold Storage Japan Inc, and Buddy Training Co, Ltd. We recruited participants through our website (Department of Preventive Services at School of Public Health, Kyoto University), our X (formerly Twitter; X Corp), posters and flyers, and the OReC—a web platform dedicated to research recruitment [23]. Further details on the recruitment procedure and materials are presented in Multimedia Appendix 1.

Sample Size

As this is a pilot RCT, a formal power-based sample size calculation for definitive hypothesis testing was not the primary objective. The main goals were to obtain estimates of effect size to inform the design of a future, larger trial. Based on recommendations for pilot study sample sizes [24], a range of 10-30 participants per group is often considered sufficient for these purposes. To ensure our sample size was within a reasonable range, we referred to previous literature [25,26] and pre-estimated the effect size to be approximately Cohen d=0.5. A sample of 10 participants per group would provide a preliminary indication of such an effect. To account for the potentially high dropout rate common in web-based intervention trials, we determined our target sample size to be 20 participants per group, for a total of 40 participants.

Randomization

Allocation

The researcher (TN) weekly sent the information (ID, sex, and age) of the registered participants to another researcher (KK) via email. He conducted a 1:1 stratified randomization using computer-generated random sequences, with sex (male or female) and age (<40 or ≥40 years) as stratifying factors. The block sizes were randomly chosen to be either 2 or 4. Allocation concealment was maintained throughout the study.

Blinding

Participants were not blinded due to the nature of the dietary intervention. The implementer of the intervention was also not blinded since it was carried out by the researcher (TN), who knew the allocation results. Although the messages sent via app were also administered by TN and not blinded, the timing of delivery and the number of characters in the messages were standardized across both groups. However, to minimize potential bias, outcome assessors and a statistician analyst were blinded, as suggested by best practices in behavioral studies [27].

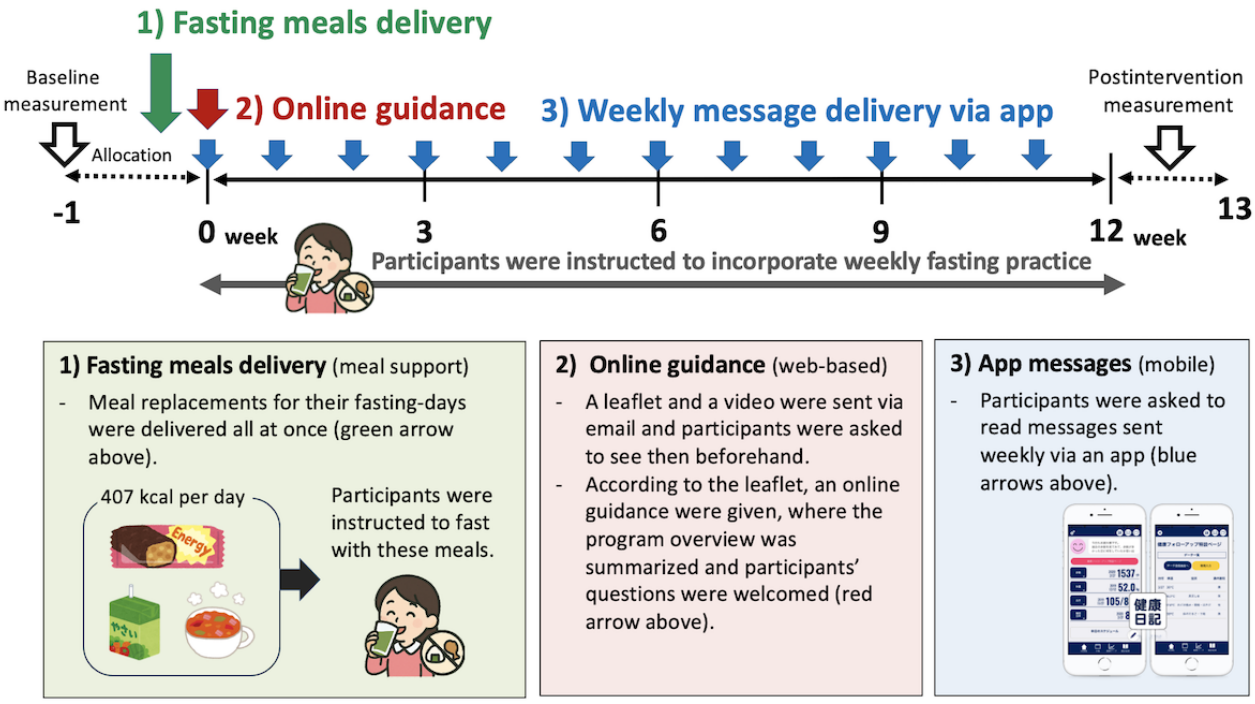
Intervention

A nonblinded researcher (TN) administered all the interventions for both groups. Detailed information on interventions was found in Multimedia Appendix 2.

Intervention Group (OIF)

Participants in the intervention group were engaged in the OIF (Figure 1). In this study, fasting was defined as consuming the provided fasting meals by 8 PM without any additional food intake. The OIF comprises three components: (1) meal support (nonweb-based component), (2) a web-based component, and (3) a mobile component. Further details are described in Multimedia Appendix 2.

Figure 1. Overview of the 12-week "Brief Online Intermittent Fasting Program (OIF).".



1. Fasting meals delivery: After the allocation, a 12-day supply of fasting meals was delivered to the intervention group in one batch. Fasting meals consisted of commercially available products that were low in energy, low in saturated fat, and plant-based and were constructed based on previous studies [28,29]. Briefly, it included 1 bottle of tomato juice, 2 bottles of fruit and vegetable juice, 2 servings of instant miso soup, 1 soy bar, and 1 multivitamin-mineral supplement, totaling 407 kcal per day, with 13% protein, 26% fat, and 61% carbohydrates. We developed this 12-week regimen in order to theoretically promote weight loss of 2.4-2.7 kg, which corresponds to a 3% loss of initial weight for a potential participant with a weight of 80 kg and 90 kg, respectively.
2. Online fasting guidance: We sent participants a leaflet detailing OIF (an A4 sheet, double-sided) via email upon allocation. They were directed to watch the prerecorded video (about 9 minutes) before the online guidance and then attended the guidance session via Zoom, which was held on their first day of the trial period and averaged 3.5 (SD 4.1) minutes per participant. They were instructed to replace one day's meals with the provided fasting meals, with a recommendation to fast on a holiday to minimize their burden.
3. Weekly message delivery via app: Participants were instructed to install an app, "Kenko-Nikki (Healthtech Laboratory [HTL], Inc)" [30], detailed in Multimedia Appendix 3. Participants received weekly messages in Japanese (mean 382, SD 126 characters) containing general health advice on diet, exercise, sleep, and encouragement and reminders for their fasting practice through the app.

Control Group—Minimal Care

Similar to the intervention group, participants were directed to read weekly messages containing general health advice on diet, exercise, and sleep. The content of these messages differed partly from those sent to the intervention group, as outlined in Multimedia Appendix 2.

Outcome Definition and Measurement

Data Collection

Overview

Sex, age, medical history, current medical visits, and medication status were collected via a web form at the time of preregistration. This information was reverified at the baseline measurement, along with identity confirmation. The following assessment data were measured at 2 time points—baseline and after 12 weeks, with the exception of adverse events and mild symptoms, adherence, and participants' feedback, which were measured once at 12 weeks, and of self-recorded weight and step counts and fasting status, which were supposed to be measured every day.

Anthropometric Measures

Body weight, fat mass, muscle mass, and basal metabolic rate (BMR) were measured by using the body composition analyzer (MC-780A-N; TANITA Corporation). Height was measured 3 times at the baseline, barefoot, using a portable height measure (Seca 213 l; Seca Japan), and the median value was used. The BMI was calculated based on these data.

Blood Pressure

Measurements of systolic blood pressure (SBP) and diastolic blood pressure (DBP) were taken twice on the upper arm opposite the dominant hand using a blood pressure monitor

(HEM–7511T; OMRON Corporation) while the participant was seated. The average value of these two measurements was used.

Blood Biomarkers

Hemoglobin A_{1c} (HbA_{1c}), triglycerides, total cholesterol, high-density-lipoprotein cholesterol (HDL-C), and low-density-lipoprotein cholesterol (LDL-C) were measured using a point-of-care biochemical analysis device (cobas b 101 plus; Roche Diagnostics KK). Blood sampling involved participants self-puncturing the fingertip of the middle finger on the opposite side of their dominant hand using a spring-loaded lancet (Nipro LS lancet 30G 1.0 mm; Nipro Corporation). Subsequently, research staff aspirated the blood onto the reagent disk for analysis in the device.

Quality of Life

Quality of life (QOL) scores were also evaluated, which were measured using an electronic scoring system for the Short-Form 12-Item Survey-version 2 (SF-12v2) Japanese version [31] (Qualitest). Participants responded to questions on the web system regarding their lives over the past month via their smartphones. These responses were then calculated through the system to obtain physical and mental 2-component summary scores, which are standardized T-scores with a mean of 50 (SD 10), where higher scores indicate better health.

To assess participants' subjective health status, responses to the first question of the SF-12v2 questionnaire, "In general, would you say your health is:," were further analyzed. Response options included "(1) Excellent," (2) Very good," (3) Good," (4) Unsatisfactory," and "(5) Poor."

Physical Activity

Participants' physical activities were self-reported using the short form of the International Physical Activity Questionnaire [32]. The questions were prepared on a Google Form, and participants were asked to respond via smartphone about their physical activity over the past month at the time of final measurement. The amount of total physical activity was calculated as metabolic equivalent of task-minutes per week (METs).

Adverse Events and Mild Symptoms

For the evaluation of the safety, the number of adverse events and the number of individuals experiencing them were investigated. Participants were instructed to consult anytime throughout the study regarding any adverse events or symptoms. In addition, they were asked at the 12-week measurement about the incidence of 20 symptoms prelisted from a prior study [33] (sleep difficulty, hunger, fatigue, headache, diarrhea, sensitivity to cold, dry mouth, back pain, bad breath, muscle pain, abdominal bloating, cravings, vertigo, blurred vision, restless leg, skin rash, nausea, palpitation, dyspepsia, and muscular cramp) via a web form. We primarily evaluated the difference in the number of these 20 events and the number of participants who reportedly experienced them. Other adverse events were also self-reported separately through a free-text web questionnaire at the measurement at 12 weeks.

Adherence

Adherence in the intervention group was defined by the number of actual fasts, based on self-report via a web questionnaire at the 12-week measurement. Adherence in the control group was defined by the read status of messages on the app, with data obtained from the server. The definition of adherence is detailed in the statistical analysis section.

Participants' Feedback

Participants' feedback on each intervention was collected through a web questionnaire at the 12-week measurement.

Self-Recorded Weight and Step Counts

Participants were directed to record their self-measured weight and step counts via the app every day throughout the trial, which was predetermined to be assessed only exploratorily.

Fasting Status

Participants in the intervention group were directed to record the following information on their fasting day: the date, the consumption status of each fasting meal (the extent to which they ate and whether they ate other food or not), and the planned date of the next fasting day. This was also to be exploratorily assessed.

Outcome

The primary outcome was the change in body weight. The changes in BMI, fat mass, muscle mass, BMR, SBP, DBP, HbA_{1c}, triglyceride, HDL-C, and LDL-C were evaluated as secondary outcomes. After formal analysis execution, the number of participants achieving a minimal clinically important change (MIC) was added as an exploratory outcome for post hoc analysis.

Statistical Analysis

Overview

The statistical analysis plan (Multimedia Appendix 4) was fixed before data locking. After finalizing the data, the precoded analysis script (R [R Foundation for Statistical Computing], version 4.3.2) was executed. Statistical estimations were presented with 95% CI. Our predetermined α level was .05. Missing data in outcomes were imputed using the Baseline Observation Carried Forward (BOCF) approach. As this was a pilot study, no formal interim analyses or prespecified stopping guidelines were planned.

Baseline Analysis and Definition of Analysis Population

Baseline characteristics were presented as the number and percentage for sex (male), while continuous variables were presented as median and IQR.

Under the intention-to-treat (ITT) principle, all the allocated participants were regarded as the primary analysis population (ITT).

The per protocol set (PPS) included individuals in the intervention group who adhered to the fasting regimen (reportedly had fasted as instructed at least 10 out of 12 times) and participants in the control group who adhered to the minimal care (had read at least 10 out of the 12 app messages, as

confirmed by server data). Additionally, within the intervention group, participants who reportedly had fasted 10 or more times out of 12 in accordance with our instructions were defined as "adherent," while those who reported fasting 12 times were categorized as "fully adherent."

Safety analyses were conducted on the completers set (CS), defined as participants who were followed up at the 12-week measurement.

In the discussion, we also presented the mean (SD) values of QOL scores for considering generalizability. Additionally, we assessed the subjective health status of participants by analyzing responses to the SF-12v2 questionnaire and comparing them with those from a similar question in a national annual survey conducted in 2022 [34]. We compare the proportions of participants who self-reported positive health status in the SF-12v2 questionnaire's first question, "In general, would you say your health is:" with those from the reference survey, which asked, "How do you rate your current health status?" with similar responses ranging from 1 to 5. In the SF-12v2, the response options were "(1) Excellent, (2) Very good, (3) Good, (4) Unsatisfactory, (5) Poor," while in the reference survey, they were "(1) Good, (2) Satisfactory, (3) Average, (4) Unsatisfactory, (5) Poor." Recognizing the differences in response expression between the 2 surveys, we focused on respondents who considered themselves healthy, combining responses categorized as 1-3. This comparison was predetermined in the protocol.

Analysis of Primary Outcome

The change in body weight was presented as the mean and SD for each group. The primary analysis was to estimate the between-group mean difference under the ITT principle, adjusted for sex, age, and baseline weight using a generalized linear regression model, presented along with a 95% CI and *P* value. The regression coefficients for each covariate were also reported with their 95% CIs. Sensitivity analyses included unadjusted comparisons of crude values (2-tailed Welch *t* test) and per-protocol analysis within the PPS (using the same adjustment as the primary analysis). Additionally, in case of missing weight data, an analysis incorporating the worst-case scenario (imputing the maximum weight gain for the intervention group and the maximum weight loss for the control group from observed data) was predetermined to be added to the sensitivity analysis.

Analysis of Secondary Outcomes

For continuous secondary outcomes, crude (unadjusted) between-group mean differences and 95% CIs were calculated using the Welch *t* test. Total physical activity was also assessed to identify within-group differences, using the Wilcoxon signed-rank test for paired comparisons. In the safety analysis, the number of adverse events and the number of individuals experiencing them were investigated, and the between-group differences in incidence proportions were analyzed by the chi-square test. Additionally, within-group differences of QOL were analyzed by the Wilcoxon signed-rank test. The proportion of follow-up rate, adherence rate, and message-read rate was presented as percentages of the ITT population.

Exploratory Analysis

Self-recorded weight and step counts were predetermined to be assessed for discussion as needed, in cases of high dropout or missing data, for instance. As a post hoc analysis for the primary outcome, the number of participants achieving a MIC was compared between groups, with *P* values reported from the Fisher exact test. In addition, we included an additional sensitivity analysis for CS, which only analyzed the participants who completed the 12-week follow-up.

Methods Alteration After Study Commencement

After the recruitment began, 3 collaborative facilities were added, and the initial eligibility criterion of having a BMI between 25 and less than 35 was modified to a BMI between 23 and less than 35. This change required a revision of the protocol and a subsequent review and reapproval by the Ethics Committee. Aspects of the statistical analysis not explicitly outlined in the original protocol were detailed in the statistical analysis plan (Multimedia Appendix 4), with its final version being fixed prior to the data lock.

Ethical Considerations

This study was conducted with the approval of the Kyoto University Graduate School and Faculty of Medicine Ethics Committee and under the authorization of the dean (approval number C1625). The study was registered in the University Hospital Medical Information Network-Clinical Trials Registry (UMIN-CTR), a clinical trial registration service that meets the standards of the International Committee of Medical Journal Editors, on February 27, 2023 (UMIN000050437). Written informed consent was obtained from all participants. To protect privacy, participant data were pseudonymized. The correspondence table linking personal information to the data was stored on a separate USB drive and kept in a locked cabinet within a locked room. This table was deleted after the analysis to prevent reidentification. Participants received a 5000 Japanese yen Amazon gift card as compensation. Other ethical considerations and the consent document were described in Multimedia Appendix 1. This trial was conducted in accordance with the CONSORT (Consolidated Standards of Reporting Trials) 2025 and CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines (Multimedia Appendix 5). There was no patient or public involvement in the design, conduct, or reporting of the study.

Results

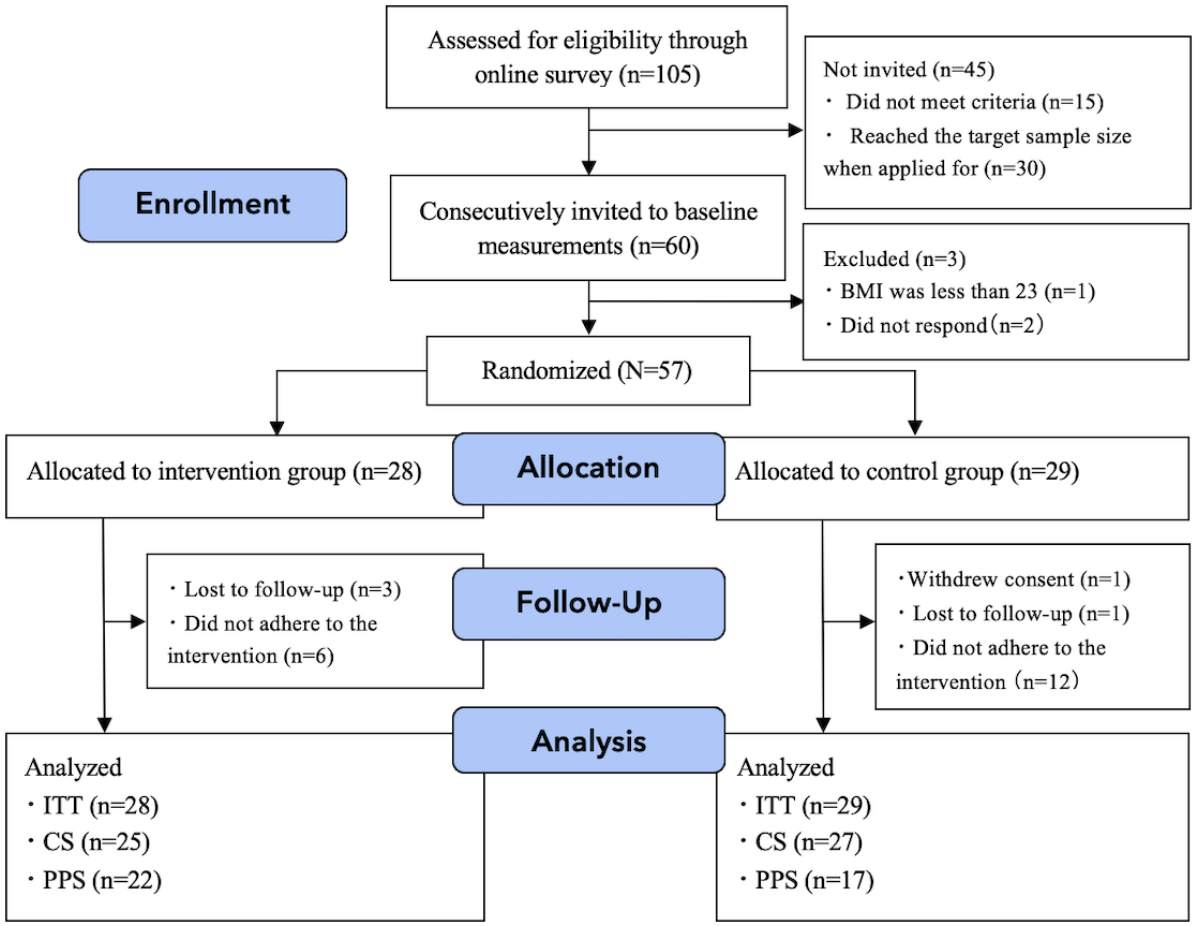
Participants Flow

Participant recruitment took place from July 20, 2023, to October 10, 2023. The follow-up measurement of all the participants ended on January 12, 2024.

Among the 60 individuals invited to the baseline measurement, 57 were eventually registered, with 28 allocated to the intervention group and 29 to the control group (Figure 2). In the intervention group, 3 participants were lost to follow-up (dropped out) due to discontinued contact. In the control group, a participant withdrew consent due to "difficulties in continuing

participation," but agreed to our using his baseline data, allowing us to include him in the ITT population. Another participant was lost to follow-up (dropped out) in the control group. Consequently, of the 57 participants in the ITT population, 89% (25/28) in the intervention group and 93% (27/29) in the control group were followed up (defined together as the CS). The PPS was defined based on adherence, with 22 (vs 17) participants in the intervention (vs control) group. Among the 28 participants in the intervention group, 22 (79%) participants were defined as "adherent" (those who fasted 10 times or more; included in PPS), while 19 (68%) were defined as "fully adherent" (those who fasted 12 times). Among the 29 participants in the control group, 17 (59%) were adherent (who had read 10 app messages or more) and thus were included in PPS. Similarly, 15 (54%) participants in the intervention group had read app messages.

Figure 2. Participants flowchart. CS: completers set; ITT: intention-to-treat; PPS: per protocol set.



Participants Characteristics

Table 1 presents the baseline characteristics of participants in each group. The proportion of men was high in both groups (19/28, 68% in the intervention group vs 19/29, 66% in the control group). The median age was 34 (IQR 29-47) vs 38 (IQR 29-52) years, the median body weight was 75.8 (IQR 68.3-80.6) vs 74.8 (IQR 69.8-81.8) kg, and the median BMI was 27.1 (IQR 24.8-29.8) vs 26.6 (IQR 24.7-28.5) for the intervention (vs control) group. The number of registered participants by facility was as follows: 22 (12 vs 10) participants at Kyoto University, 11 (7 vs 4) participants at Kusaka Hospital, 10 (4 vs 6) participants at Cold Storage Japan Inc, and 14 (5 vs 9) participants at Buddy Training Co, Ltd. Significant imbalances were not observed between the 2 groups across all characteristics.

Table 1. Baseline characteristics of the intervention and control groups (N=57)a.

Characteristic	Intervention (n=28)	Control (n=29)
Sex (male), n (%)	19 (68)	19 (66)
Age (years), median (IQR)	34 (29-47)	38 (29-52)
Weight (kg), median (IQR)	75.8 (68.3-80.6)	74.8 (69.8-81.8)
BMI (kg/cm^2), median (IQR)	27.1 (24.8-29.8)	26.6 (24.7-28.5)
Height (cm), median (IQR)	168 (163-171)	170 (163-175)
Fat mass (kg), median (IQR)	22 (18-28)	22 (18-29)
Muscle mass (kg), median (IQR)	52 (42-56)	52 (41-57)
BMR[b] (kcal), median (IQR)	1529 (1382-1682)	1563 (1369-1703)
SBP[c] (mm Hg), median (IQR)	128 (118-141)	130 (117-142)
DBP[d] (mm Hg), median (IQR)	87 (79-96)	90 (80-99)
HbA_{1c}[e] (%), median (IQR)	5.4 (5.2-5.6)	5.4 (5.2-5.5)
TG[f] (mg/dL), median (IQR)	107 (70-209)	131 (111-208)
Total-C[g] (mg/dL), median (IQR)	187 (165-212)	196 (176-220)
HDL-C[h] (mg/dL), median (IQR)	57 (45-65)	57 (48-66)
LDL-C[i] (mg/dL), median (IQR)	103 (89-122)	112 (80-128)
Total physical activity (METs[j]), median (IQR)	999 (399-1706)	1074 (827-1497)
QOL[k] score[l], median (IQR)		
Physical component summary	50 (47-52)	52 (48-54)
Mental component summary	59 (54-63)	55 (49-58)
Subjective health status[m], n (%)		
Excellent	3 (11)	3 (10)
Very good	9 (32)	5 (17)
Good	15 (54)	15 (52)
Fair	0 (0)	5 (17)
Poor	1 (4)	1 (3)
Registered facility, n (%)		
Kyoto University	12 (43)	10 (35)
Kusaka Hospital	7 (25)	4 (14)
Cold Storage Japan Inc	4 (14)	6 (21)
Buddy Training Co, Ltd.	5 (18)	9 (31)

[a]Data were presented as numbers for sex (male) and registered facilities, and as medians (IQR) for all other variables. The values were rounded for all the variables except for weight, BMI, and hemoglobin A_{1c}.

[b]BMR: basal metabolic rate.

[c]SBP: systolic blood pressure.

[d]DBP: diastolic blood pressure.

[e]HbA_{1c}: hemoglobin $A_{1c.}$

[f]TG: triglyceride.

[g]Total-C: total cholesterol.

[h]HDL-C: high-density-lipoprotein cholesterol.

[i]LDL-C: low-density-lipoprotein cholesterol.

[j]METs: metabolic equivalent of task-minutes per week.

[k]QOL: quality of life.

[l]SF-12v2 (Short-Form 12-Item Survey-version 2) Japanese version was used to obtain each 2-component summary score, which is standardized T-scores with a mean of 50 (SD 10), where higher scores indicate better health.

[m]Calculated from the responses to the first question of the SF-12v2 questionnaire, "In general, would you say your health is:" with options ranging from (1) Excellent, (2) Very good, (3) Good, (4) Unsatisfactory, and (5) Poor.

Changes in Primary Outcome

Primary Analysis

Table 2 shows the primary results of this study. After 12 weeks, the mean weight change in the intervention group was –0.9 (SD 1.9) kg, while it was 0.6 (SD 1.4) kg in the control group. The primary analysis revealed an adjusted mean difference of –1.6 (95% CI –2.5 to –0.8; *P*<.001) kg. The regression coefficients for covariates were as follows: baseline weight=–0.2 (95% CI –0.6 to 0.2) kg per 10-kg increase; sex (male)=0.5 (95% CI –0.5 to 1.4) kg; and age=–0.4 (95% CI –0.7 to 0.0) kg per 10-year increase.

Table 2. The effect of the intervention between the intervention and control groups (N=57)a.

Variable	Intervention (n=28)	Control (n=29)	Adjusted mean difference (95% CI)[b]	*P* value
Weight at baseline (kg), median (IQR)	75.8 (68.3-80.6)	74.8 (69.8-81.8)	—[c]	—
Weight at 12 weeks (kg), median (IQR)	74.5 (66.3-79.5)	74.8 (69.1-83.1)	—	—
Weight change (kg), mean (SD)	–0.9 (1.9)	0.6 (1.4)	–1.6 (–2.5 to –0.8)[b]	<.001

[a]Intention-to-treat analysis. Data were presented as median (IQR) for each measured weight and as mean (SD) for the weight change.

[b]Adjusted for stratifying factors (age and sex) and baseline weight by using the generalized linear regression model.

[c]Not applicable.

Sensitivity Analysis

Among the sensitivity analyses, the crude comparison showed a mean difference of –1.5 (95% CI –2.4 to –0.7; *P*=.001) kg, and the per-protocol analysis revealed a mean difference of –1.8 (95% CI –2.9 to –0.6; *P*=.004) kg. Additional sensitivity analysis for CS presented similar results; a mean difference of –1.8 (95% CI –2.7 to –0.9; *P*=.004) kg. In a worst-case scenario, where missing data were imputed (+3.3 kg to the weight change for the 3 missing participants in the intervention group and –4.7 kg to the weight change for the 2 missing in the control group), the adjusted mean difference was –1.0 (95% CI –2.1 to 0.2; *P*=.10) kg, and the crude mean difference was –0.9 (95% CI –2.0 to 0.3; *P*=.13) kg.

Exploratory Analysis

The post hoc analysis revealed that more participants (8/28, 29%; *P*=.01) achieved a MIC (3% reduction in initial weight) in the intervention group compared to the control group (1/29, 3%).

Changes in Secondary Outcomes

Changes in Continuous Variables

The results for secondary outcomes were presented in Table 3, showing the within-group changes and between-group differences. In line with weight changes, the BMI decreased more in the intervention group compared to the control group (mean difference=–0.4, 95% CI –0.1 to –0.8 kg). However, no significant difference was observed in fat mass changes (mean difference=–0.1, 95% CI –1.3 to 1.4 kg), whereas a decrease in muscle mass was implied (mean difference=–1.3, 95% CI –2.5 to –0.2 kg). In alignment with the changes in body composition, a decrease was suggested in the BMR (mean difference=–37.8, 95% CI –67.5 to –8.2 kcal). No significant differences were observed in other outcomes. The mean differences (95% CI) for each outcome were as follows: SBP= –4.6 (–10.6 to 1.3) mm Hg, DBP=–1.1 (–5.6 to 3.3) mm Hg, HbA_{1c}=–0.0 (–0.2 to 0.1) %, TG=–30.8 (–71.3 to 9.7) mg/dL, total cholesterol=–8.2 (–22.7 to 6.4) mg/dL, HDL-C=2.2 (–2.7 to 7.1) mg/dL, LDL-C=–4.1 (–17.1 to 8.9) mg/dL, physical QOL=0.6 (–2.6 to 3.9), mental QOL=–0.7 (–4.0 to 2.3), subjective health status=–0.2 (–0.5 to 0.3), and total physical activity=291 (–508 to 1090) METs.

Table 3. The changes in secondary outcomes from baseline to 3 months. (N=57)a.

Variable	Intervention (n=28), mean (SD)	Control (n=29), mean (SD)	Between-group difference[b], mean difference (95% CI)
BMI (kg/cm^2)	–0.3 (0.7)	0.2 (0.5)	–0.4 (–0.8 to –0.1)
Fat mass (kg)	–0.2 (1.7)	–0.3 (3.1)	0.1 (–1.3 to 1.4)
Muscle mass (kg)	–0.6 (1.0)	0.8 (2.8)	–1.3 (–2.5 to –0.2)
BMR[c] (kcal)	–17.9 (31.9)	19.9 (72.0)	–37.8 (–67.5 to –8.2)
SBP[d] (mm Hg)	–3.3 (9.4)	1.3 (12.7)	–4.6 (–10.6 to 1.3)
DBP[e] (mm Hg)	–3.6 (6.6)	–2.4 (9.8)	–1.1 (–5.6 to 3.3)
HbA_{1c}[f] (%)	–0.1 (0.2)	–0.0 (0.3)	0.0 (–0.2 to 0.1)
TG[g] (mg/dL)	1.1 (64.8)	31.9 (86.4)	–30.8 (–71.3 to 9.7)
Total-C[h] (mg/dL)	3.4 (24.4)	11.6 (30.2)	–8.2 (–22.7 to 6.4)
HDL-C[i] (mg/dL)	2.4 (7.4)	0.2 (10.8)	2.2 (–71.3 to 9.7)
LDL-C[j] (mg/dL)	0.9 (25.3)	5.0 (23.7)	–4.1 (–17.1 to 8.9)
QOL[k] score[l]			
Physical component summary	0.5 (5.8)	–0.1 (6.3)	0.6 (–2.6 to 3.9)
Mental component summary	–0.5 (6.8)	0.2 (4.6)	–0.7 (–4.0 to 2.3)
Subjective health status[m]	–0.3 (0.6)	–0.1 (0.9)	–0.2 (–0.5 to 0.3)
Total physical activity (METs[n])	350 (1679)	59 (1295)	291 (–508 to 1090)

[a]Intention-to-treat analysis. Data were presented as mean (SD) for the change in each outcome and as mean difference (95% CI) for the difference between groups in the change scores.

[b]The difference in secondary outcomes between groups was not adjusted and was presented as crude values, with 95% CI of the Welch *t* test.

[c]BMR: basal metabolic rate.

[d]SBP: systolic blood pressure.

[e]DBP: diastolic blood pressure.

[f]HbA_{1c}: hemoglobin $A_{1c.}$

[g]TG: triglyceride.

[h]Total-C: total cholesterol.

[i]HDL-C: high-density-lipoprotein cholesterol.

[j]LDL-C: low-density-lipoprotein cholesterol.

[k]QOL: quality of life.

[l]SF-12v2 (Short-Form 12-Item Survey-version 2) Japanese version was used to obtain each 2-component summary score, which is standardized T-scores with a mean of 50 (SD 10), where higher scores indicate better health.

[m]Calculated from the responses to the first question of the SF-12v2 questionnaire, “In general, would you say your health is:” with options ranging from (1) Excellent, (2) Very good, (3) Good, (4) Unsatisfactory, and (5) Poor.

[n]METs: metabolic equivalents-minutes per week.

Safety Analysis

Table 4 displays the reported number of events and the incidence—the number of participants who experienced it—for each of the 20 mild symptoms. A larger proportion of participants reported hunger in the intervention group (22/25, 89%; *P*<.001) compared to the control group (3/27, 11%). Additionally, fatigue (16/25, 64% vs 6/27, 22%; *P*=.06) and cravings (15/25, 60% vs 3/27, 11%; *P*=.01) were also more reported in the intervention group (vs control group). On the other hand, the participants in the control group reported hunger (43 vs 27) and cravings (70 vs 42) more frequently compared to the intervention group. Adverse events reported in free text included influenza-like illness (1/25 vs 3/27), sinusitis (0/25 vs 1/27), viral gastroenteritis (1/25 vs 0/27), loss of appetite (1/25 vs 3/27), and drowsiness (2/25 vs 0/27). Regarding the within-group difference in QOL scores, no significant changes were observed before and after in either group: physical QOL (intervention: mean 0.51, SD 5.84 vs control: mean –0.13, SD 6.27), and the mental QOL (mean –0.54, SD 6.84 vs mean 0.15, SD 4.60).

Table 4. Self-reported mild symptoms in the intervention and control groups (N=52)a.

Mild symptoms	Intervention (n=25)		Control (n=27)		*P* value[b]
	Events	Incidence (%)	Events	Incidence (%)	
Hunger	27	22 (88)	43	3 (11)	<.001
Fatigue	187	16 (64)	27	6 (22)	.01
Headache	125	12 (48)	74	7 (25)	.17
Cravings	42	15 (60)	70	3 (11)	.001
Diarrhea	30	5 (20)	68	7 (25)	.86
Dry mouth	118	8 (32)	68	4 (14)	.25
Back pain	66	6 (24)	21	6 (22)	>.99
Sensitivity to cold	36	5 (20)	25	6 (22)	>.99
Muscle pain	24	4 (16)	4	5 (18)	>.99
Abdominal bloating	9	4 (16)	14	5 (18)	>.99
Vertigo	17	4 (16)	9	5 (18)	>.99
Sleep difficulty	77	4 (16)	24	4 (14)	>.99
Dyspepsia	13	2 (8)	57	5 (18)	.48
Bad breath	7	4 (16)	12	2 (7)	.59
Blurred vision	6	2 (8)	0	4 (14)	.74
Nausea	0	3 (12)	13	2 (7)	.93
Muscular cramp	9	3 (12)	8	2 (7)	.93
Palpitation	1	1 (4)	7	3 (11)	.66
Skin rash	22	0 (0)	21	3 (11)	.26
Restless leg	4	2 (8)	2	0 (0)	.44

[a]Analysis population was the completers set. Participants were asked to report the number of incidences of 20 mild symptoms listed in the web questionnaire by answering 0-90 times per symptom.

[b]*P* values were presented for the incidences.

Discussion

Principal Findings

This study aimed to evaluate the effectiveness, feasibility, and safety of a 12-week OIF among adults with overweight and obesity in Japan. Overall, the program demonstrated effectiveness in promoting modest weight loss over 12 weeks, with high feasibility and safety indicated by low dropout rates and absence of serious adverse events.

The participants in the intervention group tended to lose weight (mean –0.9, SD 1.9 kg), whereas the control group, on average, gained weight (mean 0.6, SD 1.4 kg). Generally, individuals with overweight tend to continue to gain weight [35], but the result suggests this increase can be mitigated by the OIF. The adjusted mean difference was –1.6 (95% CI –2.5 to –0.8) kg, and more participants achieved a MIC in the intervention group compared to the control group (8/28 vs 1/29 participants). Among the sensitivity analyses, the per-protocol analysis showed a subtle larger effect (mean difference –1.8, 95% CI –2.9 to –0.6 kg), implying the efficacy of the OIF. The effect size calculated by the weight change (SD) was large (Cohen *d*=0.95), but the effect size based on the after-weight (SD) was small (Cohen *d*=0.19).

The dropout rate was low in the intervention group (3/28, 11%). Additionally, adherence rate exceeded our expectations, with 79% (22/28) of “adherent” and 68% (17/28) of “fully adherent,” indicating a modest level of feasibility. However, there were also concerns regarding their self-reported mild symptoms. Based on the incidence, more participants reported hunger or cravings in the intervention group, although the participants in the control group reported them more frequently in terms of the number of events they experienced. There was a slight trend that participants in the intervention group had more headaches (both in events and incidence), but the result was not statistically significant. Except for hunger, fatigue, and cravings, there were no statistically significant differences in reported symptoms between the groups.

Concerning safety, the intervention showed encouraging indications. No serious adverse events were reported, and QOL remained stable (insignificant within-group differences in QOL score between before and after, and insignificant between-group differences in change in QOL) throughout the study. These findings, coupled with the modest adherence observed, suggest that the intervention is generally safe. Nonetheless, the intervention group reported more hunger, fatigue, and cravings,

underscoring the importance of providing thorough explanations and follow-up.

Comparison to Prior Work

Many studies have demonstrated the effectiveness of web-based weight loss interventions [7,8,25]. A previous RCT among adults with overweight in Japan showed a moderate weight loss effect from a 12-week app intervention (adjusted mean difference=–1.60, 95% CI –2.83 to –0.38; *P*=.01) [26]. A recent meta-analysis indicated that web-based interventions facilitate weight loss compared to minimal care control groups (mean difference=–1.40, 95% CI –1.98 to –0.82; *P*<.001) [8]. This study observed similar effects to these. However, we initially had expected a potential weight loss of 2.7 kg through the OIF, but it was 0.9 kg, and the observed between-group difference was –1.6 kg, with both changes falling short of the expectation. A possible reason for the slight weight loss in the intervention group and the weight gain in the control group could be attributed to the timing of the study. The study was conducted from September 2023 to January 2024, and most participants completed their follow-up measurements during the Christmas and New Year holidays, which possibly contributed to the weight gain [36]. Since total physical activity did not change, another possible reason should also lie in diet. Participants in the intervention group might feel hungrier, which is consistent with the previous studies [37,38], and might have compensated for the energy reduced by fasting through consuming more food on nonfasting days. This was implied by several participants' feedback in the questionnaire. However, this is just speculative since prior works have demonstrated the lack of compensation for energy intake in IF [39-41], and this study did not collect information on a regular diet. The next study should consider comprehensive dietary assessments to elucidate these findings further.

This study showed no significant reduction in fat mass (mean difference=–0.1, 95% CI –1.3 to 1.4 kg) but a decrease in muscle mass (mean difference=–1.3, 95% CI –2.5 to –0.2 kg). However, it is natural that glycogen stored in muscles is used first during the initial phase of short-term weight loss, accompanied by a temporary decrease in water and muscle mass [42]. Thus, the observed reduction in muscle mass may not directly raise concerns. A systematic review and meta-analysis comparing conventional dietary restrictions with IF have not found evidence that IF promotes greater muscle loss [43]. Furthermore, other studies have shown that incorporating exercise during IF can increase muscle mass and exercise performance [44,45]. A recent systematic review has also concluded that combining IF with strength training can maintain lean body mass [46]. These findings suggest that the OIF does not necessarily worsen body composition, but on its own, it might not provide sufficient health benefits. Weight loss interventions should ideally include exercise therapy and comprehensive lifestyle modifications [47,48], a principle that we conclude also applies to online fasting interventions.

Strengths and Limitations

One of the strengths of this study lies in its study design. A randomized controlled design using minimal care as a control enabled us to obtain more valid estimates than study designs with single-arm or no-intervention controls. The sample size that was large for a pilot study led to demonstrating statistically significant results for weight loss, justifying conducting future trials.

In addition, we comprehensively evaluated the effects of the intervention, measuring multiple outcomes, allowing us to obtain invaluable insights into our program. For example, no observed change in physical activity implied that the OIF might facilitate weight loss through reduced energy intake.

This study has several limitations. First, we lacked important data on physical activity and dietary habits. In this study, physical activity was measured by self-report and was considered to be of low accuracy, which might prevent us from detecting potential changes. Additionally, the lack of data on a regular diet limited the interpretation of the results. In particular, evaluating why weight loss was below expectations is quite challenging. Future studies should evaluate the physical activity and dietary habits more accurately in order to distinguish the efficacy of IF from other lifestyle factors.

Second, our analysis is limited by the way of handling the missing values. We used the BOCF approach to impute missing body weight data. However, since missing data on body weight after the intervention could occur systematically, this approach may have introduced bias and compromised our conclusions. Notably, the statistically significant between-group difference in weight change diminished when considering the worst-case scenario (imputing +3.3 kg of weight change for the 3 missing participants in the intervention group and –4.7 kg for the 2 missing participants in the control group). Nevertheless, we believe that the BOCF approach is unlikely to have significantly overestimated the intervention effect. Participants in the control group were unlikely to lose weight without intervention and might even have gained weight as previously reported [35], whereas those in the intervention group were more likely to achieve weight loss. Thus, imputing the outcome weight with the baseline value is more likely to lead to diluting the weight loss effect, not overestimating it. Furthermore, our analysis demonstrated that a greater proportion of participants in the intervention group achieved a clinically meaningful weight loss (MIC; a 3% reduction in initial weight) compared to the control group, supporting the effectiveness of the intervention.

Third, as for the participants' recruitment, we used convenience sampling and cannot assume that it was a random sampling. Notably, participants included several direct acquaintances of the researchers (TN and TI). The proportion of participants considering themselves to be healthy was 88% (50/57), which is comparable to the results from the reference national survey [34], with 86.16% (91768/106507) for the total population and 89.66% (50816/56675) for those aged 20-65 years, implying that they were not far from the generally healthy population with overweight and obesity. However, our participants were exclusively Japanese, and it is unclear whether the findings would be applicable in other countries or regions with different lifestyles. Thus, the generalizability of the conclusions remains unclear. Nevertheless, we believe that our setting with multiple facilities possibly contributed to recruiting participants with a broad range of characteristics. In the future trial, we plan to

broaden the generalizability of the findings by using systematic recruitment methods and conducting the study in other countries.

Finally, we did not systematically collect data on concomitant care that participants may have received during the trial, such as seeking external dietary advice or using other commercial weight-loss programs. Therefore, we cannot completely rule out the possibility that such unmeasured care could have influenced the results.

Conclusions

The 12-week OIF can promote weight loss in apparently healthy adults with overweight and obesity; yet, it may not result in improvements in body composition. The observed low dropout rate and modest adherence, with no serious adverse events or significant changes in QOL, indicate the feasibility and safety of the program.

Acknowledgments

Special thanks to TANITA Corporation for lending the body composition analyzer at no cost. We extend our gratitude to Fujita Hideo and his coworkers at Kusaka Hospital, Daigo Goto at Cold Storage Japan Inc, and Yuzo Hoshino at Buddy Training Co, Ltd. for their collaboration. Our thanks also go to Toshiaki Furukawa, Yusuke Ogawa, and Hiroshi Okada for their invaluable advice. We are grateful to Ryuta Onodera, Daisuke Sasaki, and Nami Noda for their dedicated support as research staff. Finally, our sincere thanks to Kazuya Goto and Rina Kitabe for their thorough review of the manuscript. In preparing this manuscript, the author (TN) used the generative artificial intelligence ChatGPT-4 (OpenAI) for the English translation and proofreading of the text. The authors have carefully reviewed all content generated with this technology, making necessary modifications, and assume full responsibility for the submitted work.

Funding

This research was funded by a donation from Ritsuan Inc, which had no role in any part of the research process.

Data Availability

The datasets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: TN, TS, KK, MI
Data curation: TN, NN
Formal analysis: TN, NN
Funding acquisition: TN, TS
Investigation: TN, KK, MI
Methodology: TN, TS, KK, MI, NN, YT
Project administration: TN, TS, TI
Resources: TN, TS, YT, TI
Software: TN
Supervision: TI
Visualization: TN
Writing–original draft: TN
Writing–review & editing: TN, TS, KK, MI, NN, YT, TI

Conflicts of Interest

Ritsuan Inc, a funder of this study, did not have any role in the design, conduct, analysis, and reporting of the study. The closed, test version of the publicly available “Kenko-Nikki” app was developed and provided for this study at a low cost by HTL (Healthtech Laboratory, Inc). The principal investigator (TI) serves as an external director (but unpaid) at HTL, and uses intellectual property that he has transferred to HTL in this study. He also accepts other collaborative research projects that use the intellectual property, making him subject to auditing. TS is employed through joint research funds between HTL and Kyoto University, and uses the intellectual property mentioned above in this study. YT is also employed under the joint research funds. TN owns the Nihonsouken LLC, running dietary guidance business that involves fasting and selling health food products that are partly related to fasting. These conflicts of interest have been reported in advance to the Kyoto University Conflict of Interest Management Office and were audited by the Audit Unit of the Department of Clinical Research Facilitation, Institute for Advancement of Clinical and Translational Science at Kyoto University Hospital.

Multimedia Appendix 1
Supplementary information on participant recruitment.
[DOCX File , 1245 KB - mhealth_v14i1e58930_app1.docx]

Multimedia Appendix 2
Supplementary information on the interventions.
[DOCX File , 173 KB - mhealth_v14i1e58930_app2.docx]

Multimedia Appendix 3
Supplementary information on research app.
[PDF File (Adobe PDF File), 2948 KB - mhealth_v14i1e58930_app3.pdf]

Multimedia Appendix 4
Statistical analysis plan.
[PDF File (Adobe PDF File), 864 KB - mhealth_v14i1e58930_app4.pdf]

Multimedia Appendix 5
CONSORT 2025 checklist.
[PDF File (Adobe PDF File), 2055 KB - mhealth_v14i1e58930_app5.pdf]

References

1. Finucane MM, Stevens GA, Cowan MJ, Danaei G, Lin JK, Paciorek CJ, Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Body Mass Index). National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. Lancet 2011;377(9765):557-567 [FREE Full text] [doi: 10.1016/S0140-6736(10)62037-5] [Medline: 21295846]
2. Obesity. World Health Organization. URL: https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity [accessed 2024-02-29]
3. Arroyo-Johnson C, Mincey KD. Obesity epidemiology worldwide. Gastroenterol Clin North Am 2016;45(4):571-579 [FREE Full text] [doi: 10.1016/j.gtc.2016.07.012] [Medline: 27837773]
4. Bakaloudi DR, Barazzoni R, Bischoff SC, Breda J, Wickramasinghe K, Chourdakis M. Impact of the first COVID-19 lockdown on body weight: a combined systematic review and a meta-analysis. Clin Nutr 2022;41(12):3046-3054 [FREE Full text] [doi: 10.1016/j.clnu.2021.04.015] [Medline: 34049749]
5. Fraticelli F, Nicola MD, Vitacolonna E. A nutritional web-based approach in obesity and diabetes before and during the COVID-19 lockdown. J Telemed Telecare 2023;29(2):91-102 [FREE Full text] [doi: 10.1177/1357633X20966933] [Medline: 33081596]
6. Ammar A, Brach M, Trabelsi K. Effects of COVID-19 home confinement on eating behaviour and physical activity: results of the ECLB-COVID19 international online survey. Nutrients 2020;12:1583. [doi: 10.37473/dac/10.1101/2020.05.04.20072447]
7. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
8. Sorgente A, Pietrabissa G, Manzoni GM, Re F, Simpson S, Perona S, et al. Web-based interventions for weight loss or weight loss maintenance in overweight and obese people: a systematic review of systematic reviews. J Med Internet Res 2017;19(6):e229 [FREE Full text] [doi: 10.2196/jmir.6972] [Medline: 28652225]
9. Akers JD, Estabrooks PA, Davy BM. Translational research: bridging the gap between long-term weight loss maintenance research and practice. J Am Diet Assoc 2010;110(10):1511-22, 1522.e1 [FREE Full text] [doi: 10.1016/j.jada.2010.07.005] [Medline: 20869490]
10. Ye Y, Zhang M, Lin Z, Tang L. Is intermittent fasting better than continuous energy restriction for adults with overweight and obesity? Diabetes Metab Syndr Obes 2022;15:2813-2826 [FREE Full text] [doi: 10.2147/DMSO.S376409] [Medline: 36134390]
11. Anton SD, Moehl K, Donahoo WT, Marosi K, Lee SA, Mainous AG, et al. Flipping the metabolic switch: understanding and applying the health benefits of fasting. Obesity (Silver Spring) 2018;26(2):254-268 [FREE Full text] [doi: 10.1002/oby.22065] [Medline: 29086496]
12. Allaf M, Elghazaly H, Mohamed O, Fareen MFK, Zaman S, Salmasi AM, et al. Intermittent fasting for the prevention of cardiovascular disease. Cochrane Database Syst Rev 2021;1(1):CD013496 [FREE Full text] [doi: 10.1002/14651858.CD013496.pub2] [Medline: 33512717]
13. Morales-Suarez-Varela M, Collado Sánchez E, Peraita-Costa I, Llopis-Morales A, Soriano JM. Intermittent fasting and the possible benefits in obesity, diabetes, and multiple sclerosis: a systematic review of randomized clinical trials. Nutrients 2021;13(9):3179 [FREE Full text] [doi: 10.3390/nu13093179] [Medline: 34579056]
14. de Cabo R, Mattson MP. Effects of intermittent fasting on health, aging, and disease. N Engl J Med 2019;381(26):2541-2551. [doi: 10.1056/NEJMra1905136] [Medline: 31881139]

15. Klempel MC, Kroeger CM, Bhutani S, Trepanowski JF, Varady KA. Intermittent fasting combined with calorie restriction is effective for weight loss and cardio-protection in obese women. Nutr J 2012;11:98 [FREE Full text] [doi: 10.1186/1475-2891-11-98] [Medline: 23171320]
16. Kroeger CM, Klempel MC, Bhutani S, Trepanowski JF, Tangney CC, Varady KA. Improvement in coronary heart disease risk factors during an intermittent fasting/calorie restriction regimen: relationship to adipokine modulations. Nutr Metab (Lond) 2012;9(1):98 [FREE Full text] [doi: 10.1186/1743-7075-9-98] [Medline: 23113919]
17. Di Francesco A, Di Germanio C, Bernier M, de Cabo R. A time to fast. Science 2018;362(6416):770-775 [FREE Full text] [doi: 10.1126/science.aau2095] [Medline: 30442801]
18. Patterson RE, Sears DD. Metabolic effects of intermittent fasting. Annu Rev Nutr 2017;37:371-393 [FREE Full text] [doi: 10.1146/annurev-nutr-071816-064634] [Medline: 28715993]
19. Jamshed H, Beyl R, Della Manna D, Yang E, Ravussin E, Peterson C. Early time-restricted feeding improves 24-Hour glucose levels and affects markers of the circadian clock, aging, and autophagy in humans. Nutrients 2019;11(6):1234 [FREE Full text] [doi: 10.3390/nu11061234] [Medline: 31151228]
20. Manoogian EN, Panda S. Circadian rhythms, time-restricted feeding, and healthy aging. Ageing Res Rev 2017;39:59-67 [FREE Full text] [doi: 10.1016/j.arr.2016.12.006] [Medline: 28017879]
21. Wilhelmi de Toledo F, Grundler F, Sirtori CR, Ruscica M. Unravelling the health effects of fasting: a long road from obesity treatment to healthy life span increase and improved cognition. Ann Med 2020;52(5):147-161 [FREE Full text] [doi: 10.1080/07853890.2020.1770849] [Medline: 32519900]
22. Obesity in adults: prevention and lifestyle weight management programmes. National Institute for Health and Care Excellence. 2016. URL: https://www.nice.org.uk/guidance/qs111 [accessed 2023-07-31]
23. Open Research Cafe. URL: https://orec.space [accessed 2023-11-29]
24. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res 2016;25(3):1057-1073 [FREE Full text] [doi: 10.1177/0962280215588241] [Medline: 26092476]
25. Muramoto A, Matsushita M, Kato A, Yamamoto N, Koike G, Nakamura M, et al. Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes Res Clin Pract 2014;8(5):e466-e475. [doi: 10.1016/j.orcp.2013.10.003] [Medline: 25263836]
26. Nakata Y, Sasai H, Gosho M, Kobayashi H, Shi Y, Ohigashi T, et al. A smartphone healthcare application, CALO mama Plus, to promote weight loss: a randomized controlled trial. Nutrients 2022;14(21):4608 [FREE Full text] [doi: 10.3390/nu14214608] [Medline: 36364870]
27. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 2012;344:e1119. [doi: 10.1136/bmj.e1119] [Medline: 22371859]
28. Wei M, Brandhorst S, Shelehchi M, Mirzaei H, Cheng CW, Budniak J, et al. Fasting-mimicking diet and markers/risk factors for aging, diabetes, cancer, and cardiovascular disease. Sci Transl Med 2017;9(377):eaai8700 [FREE Full text] [doi: 10.1126/scitranslmed.aai8700] [Medline: 28202779]
29. Wilhelmi de Toledo F, Buchinger A, Burggrabe H, Hölz G, Kuhn C, Lischka E, Medical Association for FastingNutrition (Ärztegesellschaft für Heilfasten und Ernährung, ÄGHE. Fasting therapy - an expert panel update of the 2002 consensus guidelines. Forsch Komplementmed 2013;20(6):434-443. [doi: 10.1159/000357602] [Medline: 24434758]
30. "Health Diary" is a free app that records your daily health status. Healthtech.Lab Inc. URL: https://htech-lab.co.jp/products/kenkounikki.html [accessed 2024-02-21]
31. Ware J, Kosinski M, Keller SD. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
32. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
33. Wilhelmi de Toledo F, Grundler F, Bergouignan A, Drinda S, Michalsen A. Safety, health improvement and well-being during a 4 to 21-day fasting period in an observational study including 1422 subjects. PLoS One 2019;14(1):e0209353 [FREE Full text] [doi: 10.1371/journal.pone.0209353] [Medline: 30601864]
34. Overview of the 2022 (Reiwa 4) comprehensive survey on living conditions of the people. Ministry of Health, Labour and Welfare. 2022. URL: https://www.mhlw.go.jp/toukei/saikin/hw/k-tyosa/k-tyosa22/index.html [accessed 2024-02-25]
35. Katsoulis M, Lai AG, Diaz-Ordaz K, Gomes M, Pasea L, Banerjee A, et al. Identifying adults at high-risk for change in weight and BMI in England: a longitudinal, large-scale, population-based cohort study using electronic health records. Lancet Diabetes Endocrinol 2021;9(10):681-694 [FREE Full text] [doi: 10.1016/S2213-8587(21)00207-2] [Medline: 34481555]
36. Helander EE, Wansink B, Chieh A. Weight gain over the holidays in three countries. N Engl J Med 2016;375(12):1200-1202. [doi: 10.1056/NEJMc1602012] [Medline: 27653588]

37. Sundfør TM, Svendsen M, Tonstad S. Effect of intermittent versus continuous energy restriction on weight loss, maintenance and cardiometabolic risk: a randomized 1-year trial. Nutr Metab Cardiovasc Dis 2018;28(7):698-706. [doi: 10.1016/j.numecd.2018.03.009] [Medline: 29778565]
38. Heilbronn LK, Smith SR, Martin CK, Anton SD, Ravussin E. Alternate-day fasting in nonobese subjects: effects on body weight, body composition, and energy metabolism. Am J Clin Nutr 2005;81(1):69-73 [FREE Full text] [doi: 10.1093/ajcn/81.1.69] [Medline: 15640462]
39. Huang A, Henderson G, Profeta A, Pfeiffer M, Feinstein LH, deLahunta M, et al. Lack of compensation of energy intake explains the success of alternate day feeding to produce weight loss. Physiol Behav 2023;263:114128. [doi: 10.1016/j.physbeh.2023.114128] [Medline: 36805441]
40. Varady KA, Bhutani S, Klempel MC, Kroeger CM, Trepanowski JF, Haus JM, et al. Alternate day fasting for weight loss in normal weight and overweight subjects: a randomized controlled trial. Nutr J 2013;12(1):146 [FREE Full text] [doi: 10.1186/1475-2891-12-146] [Medline: 24215592]
41. Ravussin E, Beyl RA, Poggiogalle E, Hsia DS, Peterson CM. Early time-restricted feeding reduces appetite and increases fat oxidation but does not affect energy expenditure in humans. Obesity (Silver Spring) 2019;27(8):1244-1254 [FREE Full text] [doi: 10.1002/oby.22518] [Medline: 31339000]
42. Denke MA. Metabolic effects of high-protein, low-carbohydrate diets. Am J Cardiol 2001;88(1):59-61. [doi: 10.1016/s0002-9149(01)01586-7] [Medline: 11423059]
43. Enríquez Guerrero A, San Mauro Martín I, Garicano Vilar E, Camina Martín MA. Effectiveness of an intermittent fasting diet versus continuous energy restriction on anthropometric measurements, body composition and lipid profile in overweight and obese adults: a meta-analysis. Eur J Clin Nutr 2021;75(7):1024-1039. [doi: 10.1038/s41430-020-00821-1] [Medline: 33293678]
44. Keenan SJ, Cooke MB, Hassan EB, Chen WS, Sullivan J, Wu SX, et al. Intermittent fasting and continuous energy restriction result in similar changes in body composition and muscle strength when combined with a 12 week resistance training program. Eur J Nutr 2022;61(4):2183-2199 [FREE Full text] [doi: 10.1007/s00394-022-02804-3] [Medline: 35084574]
45. Martínez-Rodríguez A, Rubio-Arias JA, García-De Frutos JM, Vicente-Martínez M, Gunnarsson TP. Effect of high-intensity interval training and intermittent fasting on body composition and physical performance in active women. Int J Environ Res Public Health 2021;18(12):6431 [FREE Full text] [doi: 10.3390/ijerph18126431] [Medline: 34198554]
46. Keenan S, Cooke MB, Belski R. The effects of intermittent fasting combined with resistance training on lean body mass: a systematic review of human studies. Nutrients 2020;12(8):2349 [FREE Full text] [doi: 10.3390/nu12082349] [Medline: 32781538]
47. Hall ME, Cohen JB, Ard JD, Egan BM, Hall JE, Lavie CJ, American Heart Association Council on Hypertension; Council on Arteriosclerosis, ThrombosisVascular Biology; Council on LifestyleCardiometabolic Health;Stroke Council. Weight-loss strategies for prevention and treatment of hypertension: a scientific statement from the American heart association. Hypertension 2021;78(5):e38-e50. [doi: 10.1161/HYP.0000000000000202] [Medline: 34538096]
48. Wharton S, Lau DC, Vallis M, Sharma AM, Biertho L, Campbell-Scherer D, et al. Obesity in adults: a clinical practice guideline. CMAJ 2020;192(31):E875-E891 [FREE Full text] [doi: 10.1503/cmaj.191707] [Medline: 32753461]

Abbreviations

BMR: basal metabolic rate
BOCF: Baseline Observation Carried Forward
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Applications and Online Telehealth
CS: completers set
DBP: diastolic blood pressure
HbA1c: hemoglobin A1c
HDL-C: high-density-lipoprotein cholesterol
HTL: Healthtech Laboratory, Inc
IF: intermittent fasting
ITT: intention-to-treat
LDL-C: low-density-lipoprotein cholesterol
METs: metabolic equivalent of task-minutes per week
MIC: minimal clinically important change
OIF: Brief Online Intermittent Fasting Program
PPS: per protocol set
QOL: quality of life
RCT: randomized controlled trial
SBP: systolic blood pressure

SF-12v2: Short-Form 12-Item Survey-version 2
UMIN-CTR: University Hospital Medical Information Network-Clinical Trials Registry

Edited by L Buis; submitted 03.Apr.2024; peer-reviewed by J Zheng, J Mistry, R Bipat; comments to author 17.Sep.2024; revised version received 31.Dec.2024; accepted 30.Sep.2025; published 26.Jan.2026.

Please cite as:
Noda T, Shimamoto T, Kiyohara K, Imanaka M, Nishioka N, Tateyama Y, Iwami T
A Brief Web-Based and Mobile Intervention of Intermittent Fasting With Meal Support for Weight Loss Among Adults With Overweight and Obesity in Japan: Pilot Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e58930
URL: https://mhealth.jmir.org/2026/1/e58930
doi:10.2196/58930
PMID:

Original Paper

Effectiveness of Smartphone-Based Dyadic Interventions to Increase Physical Activity in Romantic Couples: Microrandomized Trial

Patrick Stefan Höhener[1], PhD; Robert Tobias[1], PhD, ING; James Martin Allen[1,2], PhD; Pascal Küng[1], PhD; Urte Scholz[1], Prof Dr

[1]Department of Psychology, University of Zurich, Zurich, Switzerland

[2]School of Health & Wellbeing, University of Glasgow, Glasgow, United Kingdom

Corresponding Author:
Patrick Stefan Höhener, PhD
Department of Psychology
University of Zurich
Binzmühlestrasse 14
8050
Zurich
Switzerland
Phone: 41 44 635 72 45
Email: patrick.hoehener@psychologie.uzh.ch

Abstract

Background: Social exchange processes, such as social support and social control, can promote health behavior change. However, these processes are often neglected when studying health behavior change and designing interventions. Intervening on these social exchange processes using dyadic interventions may provide a promising approach to promote health behaviors.

Objective: This study aimed to investigate the effects of dyadic interventions to increase moderate-to-vigorous physical activity (MVPA) in romantic couples. Furthermore, we explored how the target, type, and timing of the interventions affect their effectiveness.

Methods: In total, 38 romantic couples (mean age 34.01, SD 11.03 y) were recruited through online advertisements and participated in a smartphone-based microrandomized trial over 55 days consisting of control and intervention phases. The fully automated dyadic interventions included a one-time psychoeducation intervention, weekly dyadic and collaborative planning, and dyadic just-in-time adaptive interventions (JITAIs). MVPA was measured through daily diaries and wrist-worn accelerometers. We used multilevel modeling to estimate the effect of the intervention phase and weighted and centered estimation for microrandomized trials to estimate the treatment effects of dyadic and collaborative planning, as well as the dyadic JITAIs.

Results: Participants indicated higher device-based (b=5.88, SE=3.04, t_{3665}=1.93; *P*=.03) and self-reported (b=8.26, SE=3.88, t_{3904}=2.13; *P*=.01) MVPA during the intervention phase compared with the control phase. Dyadic and collaborative planning did not increase device-based (b=6.31, SE=5.18; *P*=.12) but only self-reported (b=14.25, SE=5.16; *P*=.005) MVPA. However, the effects of the 2 kinds of planning on self-reported MVPA disappeared when additional covariates were included (b=0.14, SE=3.32; *P*=.48). Furthermore, the dyadic JITAIs targeting both the actor and the partner increased device-based (actor: b=11.17, SE=3.18; *P*<.001; partner: b=7.23, SE=3.60; *P*=.03) and self-reported (actor: b=17.34, SE=3.65; *P*<.001; partner: b=11.82, SE=4.10; *P*<.001) MVPA. However, the effects of the dyadic JITAIs targeting the actor disappeared for self-reported MVPA (b=2.20, SE=3.22; *P*=0.25) when additional covariates were included. Exploratory analyses revealed that different types and timings of dyadic JITAIs were differentially effective.

Conclusions: This study demonstrated the promising effects of dyadic interventions to promote MVPA and highlighted the importance of the target, type, and timing of the interventions. Further research should investigate the mechanisms underlying the effects of dyadic interventions on health behaviors.

Trial Registration: ISRCTN registry ISRCTN15673058; https://www.isrctn.com/ISRCTN15673058

(JMIR Mhealth Uhealth 2026;14:e67136) doi:10.2196/67136

KEYWORDS

health behavior change; just-in-time adaptive interventions; dyadic interventions; romantic couples; physical activity; microrandomized trial; mobile phone

Introduction

Background

Regular engagement in physical activity is essential for maintaining overall health and reducing the risk of chronic and noncommunicable diseases, including heart disease, stroke, diabetes, and cancer [1]. The World Health Organization recommends at least 75 to 150 minutes of moderate-to-vigorous physical activity (MVPA) per week. Even though most people know of its importance and intent to engage in MVPA, many struggle to engage in the recommended amount of MVPA [2]. For example, in Switzerland, where the study took place, about 24% of adults do not meet these recommendations [3].

Theoretical and empirical research has been conducted to support people in increasing their MVPA [4]. This research has primarily focused on behavior change on the individual level and neglected social relationships and social processes [5]. However, since people are embedded in social networks, health behavior change often occurs in a social context. These social networks promote interdependence among people, and especially close partners can influence each other's health behaviors [6,7]. Thus, involving close social relationships in the behavior change process and implementing dyadic interventions provides a promising approach to promoting health behavior change in people with low physical activity [8].

Dyadic Interventions for Health Behavior Change

Implementing dyadic interventions makes it possible to capitalize on the interpersonal resources of both partners, encouraging mutual support and enhancing investment in the partners' behavior change, thereby facilitating behavior change [9-11]. The empirical evidence regarding the effectiveness of dyadic interventions is promising but mixed. On the one hand, some studies have shown that dyadic interventions are effective in increasing physical activity [12] and are more effective than control interventions, including equivalent interventions targeting individuals [9]. On the other hand, systematic reviews found mixed evidence for improved physical health [13], and they did not demonstrate superior effectiveness in increasing physical activity compared with comparison groups involving individual or control interventions [11].

Dyadic interventions may be categorized along the continuum of dyadic interventions into different prototypes of dyadic interventions depending on the degree of partner involvement (eg, cross-over and joint interventions) [10]. Cross-over interventions prompt interactions between the individuals by explicitly instructing one dyad member to engage with or clearly refer to the other dyad member, but the intervention is not administered to both. Joint interventions actively involve both dyad members and consider the dyad as a unit, making it impossible to be delivered to only 1 dyad member present. Cross-over and joint interventions may work through different mechanisms and thus may differ in their effectiveness. However, little research has investigated the differences in their effectiveness, calling for a more nuanced examination of these 2 prototypes of dyadic interventions [10].

Dyadic interventions may work by applying different dyadic behavior change techniques (DBCTs) [10] that explicitly elicit various social exchange processes influencing health behavior change [14]. Social exchange processes are interactions between dyad members that affect one or both individuals' behaviors, emotions, or cognitions [15]. There are various social exchange processes, such as providing social support, exerting social control, or collaboratively establishing goals and plans. Social support refers to providing psychological, material, or informational resources intended to benefit the ability to cope with stressors, solve problems, and pursue life opportunities [16,17]. One common distinction of social support is between emotional and instrumental support. Emotional support concerns the emotional well-being of the recipient [18]. Regarding physical activity, this form of support may include encouragement and comfort, guidance, and help to stay motivated and committed. Instrumental support is defined as providing the recipient with practical help and resources, such as advice or assistance [18]. Regarding physical activity, this form of support could include providing help while exercising or taking over daily chores to free up time for the partner to engage in MVPA. Empirical evidence showed that emotional and instrumental support can be associated with higher levels of physical activity, but there is also substantial heterogeneity in the associations (eg, studies by Kouvonen et al [19], Rackow et al [20], and Scarapicchia et al [21]).

Another social exchange process is social control, which refers to a deliberate and intentional attempt to change what another person thinks, feels, or does toward an outcome desired by the person exerting control [22]. Social control can be categorized into positive or negative social control based on the means used to influence behavior [23]. Positive social control includes strategies such as persuasion, modeling, and discussions. Negative social control includes, for example, coercion, social pressure, elicitation of negative emotions, and withdrawal. These strategies may differ in effectiveness when used to promote health behaviors in the partner. A meta-analysis has found that positive control was associated with better health behaviors, whereas negative control was related to worse health behaviors [22]. Furthermore, negative control relates to increased reactance-related responses and resistance to change, rendering it ineffective in encouraging health behavior change [7]. So far, there is almost no research on the effectiveness of interventions promoting positive social control and reducing negative social control. This gap will be addressed in this study.

While social support and positive social control can enhance health behaviors, poorly executed or nonresponsive support may yield adverse outcomes, including diminished well-being, heightened stress, and decreased performance [24]. Therefore, when deploying dyadic interventions that leverage social

exchange processes, it seems crucial for partners to have the knowledge and skills to offer effective support. However, little research has included information on skillful social support in interventions.

Social exchange processes can also play a role in goal setting and action planning. Goal setting refers to defining behaviors or states one wants to accomplish in the future [25]. Action planning refers to linking behaviors to specific cues by specifying when, where, and how to act [26]. Action planning can be augmented by coping planning, a barrier-focused self-regulation strategy in which an association between anticipated barriers and suitable solutions to overcome these barriers is made [26]. Combining action planning with coping planning can have additive and synergistic effects to promote health behaviors [27]. Furthermore, setting goals and planning physical activities with a partner may facilitate adherence to these goals and plans [5,28]. There are 2 forms of planning in the dyadic context. Dyadic planning involves creating plans with a partner on when, where, and how one partner will implement a behavior [29]. Collaborative planning entails creating joint plans with a partner on when, where, and how both partners will engage in a behavior [30]. There is mixed evidence for the effectiveness of dyadic planning in promoting health behaviors. While some studies have found positive associations between dyadic planning and goal progress [31] and plan enactment [32], other studies have not supported its effectiveness [33,34]. However, empirical evidence generally supports the effectiveness of collaborative planning to foster health behaviors [28,35,36].

Just-in-Time Adaptive Interventions

Technological advancements and the widespread use of devices, such as smartphones and smartwatches, have opened up new possibilities for enhancing the design and implementation of interventions [37]. These technologies allow for tracking health behaviors and context variables, which provide information about the individual's current state. This information can be integrated with theoretical and empirical knowledge to tailor the intervention content to individual needs and deliver interventions at the right time. Such interventions are known as just-in-time adaptive interventions (JITAIs) [38]. A meta-analytical review found moderate to large effects of JITAIs compared with waitlist-control conditions and non-JITAI treatments to improve health behaviors [39]. However, regarding physical activity and sedentary behavior, the evidence supporting the effectiveness of JITAIs is mixed [37]. However, so far, only JITAIs aimed at individuals have been implemented. Combining DBCTs with JITAIs and intervening on a dyadic level may provide a promising approach to increasing their effectiveness in promoting MVPA.

This Study

This study aimed to examine the effectiveness of dyadic interventions in promoting MVPA in romantic couples. Specifically, we investigated the effectiveness of dyadic interventions containing various DBCTs that elicit social exchange processes for increasing MVPA in romantic couples. We implemented an intensive-longitudinal study over 55 days consisting of phases without dyadic interventions (ie, control phase) and with dyadic interventions, including a one-time psychoeducation intervention on skilled support, weekly planning interventions, and dyadic JITAIs (ie, intervention phase). In the psychoeducation intervention on skilled support, participants learned how to support their partner appropriately to increase their MVPA. In the planning interventions, the participants set a goal and planned the physical activities for the upcoming week. Finally, there were dyadic JITAIs targeting various social exchange processes to increase MVPA.

To investigate the effectiveness of the dyadic interventions, we compared the MVPA during the intervention phase with the control phase. To isolate the effects of the different dyadic interventions, we investigated the effectiveness of the planning interventions and the dyadic JITAIs separately. Note that we did not examine the effect of the skilled support intervention since this was a one-time intervention taking place right after the baseline week and is, therefore, highly correlated with the intervention phase.

We proposed the following hypotheses, all of which were preregistered on Open Science Framework (OSF):

- Hypothesis 1: Couples show higher MVPA (device-based and self-reported) during the intervention phase than during the control phase.
- Hypothesis 2: On days for which participants had planned to be physically active during the planning intervention, they engage in more MVPA (device-based and self-reported) compared with days for which they had not planned any physical activity.
- Hypothesis 3a: Participants indicate higher MVPA (device-based and self-reported) on days when a dyadic JITAI targeted the actor's MVPA compared with days without a dyadic JITAI targeting the actor.
- Hypothesis 3b: Participants indicate higher MVPA (device-based and self-reported) on days when a dyadic JITAI targeted the partner compared with days without a dyadic JITAI targeting the partner.

Additionally, we conducted 3 exploratory analyses. First, we explored the effect of different types of dyadic JITAIs by differentiating between cross-over and joint interventions. Second, we explored the effectiveness of the dyadic JITAIs depending on their timing (ie, when they were sent). There were dyadic JITAIs sent before the planning intervention, before a planned activity, and in the evening. These 3 timings target different aspects of the behavior change process, which may vary in their effectiveness. Finally, we explored how long the effects of the dyadic JITAIs last. Dyadic JITAIs may make lasting changes in the interaction patterns between the partners, allowing their effects to be maintained over time. Thus, these changes in the social exchange processes may go beyond the specific day targeted by the dyadic JITAIs, facilitating engagement in MVPA at a later occasion.

Methods

Funding and Preregistration

This study is part of the "Time and Ties: Dynamic modelling of temporal patterns in dyadic health behaviour change" project

funded by the Swiss National Science Foundation (grant 10001C_197471 / 1). A comprehensive description of the project can be found on the OSF page of the Time and Ties project [40]. The preregistration, including hypotheses, inclusion and exclusion criteria, and planned primary analyses, can be found on the OSF page of the study [41]. The study differs from the preregistration in that we conducted the dyadic analyses for indistinguishable instead of distinguishable dyads. This decision was made to be inclusive of same-gender couples. The completed CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist can be found in Multimedia Appendix 1.

Ethical Considerations

The study received approval from the Ethics Committee of the Faculty of Arts and Social Sciences of the University of Zurich (approval 21.9.11). All participants provided online informed consent using a checkbox before enrolling in the study and were free to withdraw at any time (the informed consent is provided in Multimedia Appendix 2). All collected data are stored in anonymized form on the server of the University of Zurich and are only accessible to the project team. All participants received 150 Swiss Francs (approximately US $189) for their complete participation.

Participants

Data collection took place from June to October 2022. Romantic couples were recruited from the Swiss population through online advertisements on Facebook (Meta) and Instagram (Meta). Inclusion criteria were (1) both partners must be at least 18 years old, (2) they must be in a romantic relationship with each other and live together, (3) both partners must be physically active for less than 150 minutes per week, (4) both partners must have the intention to be more physically active, (5) both partners must not have severe health conditions preventing physical activity, (6) both partners must have a smartphone that they use regularly and have sufficient literacy to operate it independently, and (7) both partners must speak German fluently.

Descriptive statistics of the sample are presented in Table 1.

Table 1. Descriptive statistics of the sample characteristics (76 participants; 38 couples).

Variable	Value
Gender, n (%)	
Women	39 (51)
Men	37 (49)
Education, n (%)	
Vocational education	8 (11)
High school diploma	24 (32)
Bachelor's degree	23 (30)
Master's degree	21 (28)
Married, n (%)	26 (34)
Having children, n (%)	21 (28)
Age (y), mean (SD; range)	34.01 (11.03; 19-60)
BMI (kg/m^2), mean (SD; range)	24.94 (4.11; 16.37-33.95)
Relationship duration (y), mean (SD; range)	9.23 (9.09; 0.58-36.00)
Cohabitation duration (y), mean (SD; range)	7.53 (9.20; 0.25-33.00)
Age of children (n=21; y), mean (SD; range)	16.16 (9.24; 4-32)

Study Design

We conducted an intensive-longitudinal microrandomized trial over 55 days. This study was the first empirical study within the larger Time and Ties project and aimed to inform a computational modeling approach on dyadic interventions as well as to test and improve the smartphone-based data gathering and intervention provision. Thus, the study was not powered for advanced statistical analyses of dyadic intervention effects. From this perspective, it can be considered a pilot trial. An overview of the larger Time and Ties project, including this study, can be found on OSF. The study was smartphone-based, using a self-developed app in collaboration with a company specializing in app development (Multimedia Appendix 3). Participants read the information about the study and provided written informed consent before the start of the study. They were informed that various interventions would be implemented during the study. After registration, the participants received a link to download and sign in to the app. Throughout the study, participants completed daily diaries and wore wrist-worn accelerometers. The study consisted of phases without dyadic interventions (ie, control phase) and with dyadic interventions, including a 1-time psychoeducation intervention about skilled support, weekly planning interventions, and dyadic JITAIs (ie, intervention phase). Random allocation was implemented using a computer program that assigned couples to the intervention conditions characterized with different intervention phases, while ensuring equal sample sizes across groups (ie, blocked

randomization). Each group started with a 1-week control phase to establish a baseline. Afterwards, group A received a 7-week intervention phase. Group B received a 3-week intervention phase followed by a 4-week control phase. Group C received a 4-week control phase followed by a 3-week intervention phase at the end of the study (Figure S1 in Multimedia Appendix 4). During the intervention phase, various dyadic interventions were implemented on designated days, but there were also randomly assigned control days without any interventions. Participants were aware that there were interventions during the study, but they did not know which intervention conditions there were. At the end of the study, participants completed a questionnaire similar to the baseline questionnaire, where they also had the opportunity to provide qualitative feedback. All questionnaires and interventions were sent using a smartphone app designed for this study.

Interventions

Skilled Support Intervention

The psychoeducation intervention about the principle of skilled support took place on the first Saturday after the 1-week control phase. During this intervention, couples learned about different types of social support and social control, the importance of the right timing and equity of support, and how to support each other effectively to help engage in MVPA (DBCTs: "one partner receives education for supporting the other partner," which was applied by both partners) [24]. Furthermore, partners discussed and reported how and when they would like to be supported and how they communicate about social support and social control (DBCTs: "one partner identifies preferred support strategies for the other partner," "one partner gives feedback on support provision of the other partner," and "one partner practices communication skills for health behavior of the other partner;" all of which were applied by both partners).

Planning Interventions

The planning interventions occurred each Sunday during the intervention phase and consisted of 3 parts. First, the couple set their weekly goal by indicating the desired duration of MVPA they aimed to achieve in the upcoming week (DBCTs: "the couple sets a goal for the couple" and "the couple commits to a goal of the couple") [42]. Second, they planned physical activities for the upcoming week. Thereby, they could either plan dyadically (ie, plan together activities they want to do individually) [29] or collaboratively (ie, plan together activities they want to do together) [30]. They planned the activity, timing, location, duration, and whether they would participate individually or together in the planned activity (DBCTs: "the couple plans for one partner" and "the couple plans for the couple"). Third, they engaged in coping planning by anticipating potential barriers that may prevent them from engaging in the planned activities and discussing solutions to overcome them (DBCTs: "the couple creates a coping plan for one partner" and "the couple creates a coping plan for the couple") [26].

Dyadic JITAIs

The dyadic JITAIs were designed in accordance with the recommendations proposed by Nahum-Shani et al [38]. The JITAIs targeted various social exchange processes within the couple that were hypothesized to promote engagement in MVPA (a complete list of all the DBCTs that were implemented can be found on OSF). Note that the target of the JITAI is not necessarily the person who receives the JITAI. For example, a JITAI prompting to provide emotional support to engage in MVPA may be sent to one partner (ie, execution level) but targeting the other (ie, target level) [10]. Depending on the content of the JITAIs, they could target either one or both partners' MVPA. However, due to the interdependence between the partners, the effects may spill over to the nontargeted partner [43]. Thus, this study included JITAIs targeting the actor's MVPA and JITAIs targeting the partner's MVPA. Note that both JITAIs targeting the actor and the partner may be present on the same day (ie, through joint or 2 cross-over interventions targeting different partners).

There were 2 types of dyadic JITAIs: cross-over and joint JITAIs [10]. Cross-over JITAIs were sent to one partner and included instructions to interact with the other partner (eg, provide emotional support to the partner). Joint JITAIs were sent to both partners and included instructions to both partners to actively interact with each other (eg, jointly create a list of advantages of physical activity). The JITAIs were sent to the participants at a specified time, hypothesized to be the most appropriate for the JITAI content. There were three timings when the JITAIs could be triggered: (1) before the planning intervention, (2) before a planned activity, and (3) in the evening. The dyadic JITAIs before the planning intervention targeted the social exchange processes during the planning; those before the planned activity targeted the social exchange processes (eg, social support and social control) before and during the activity; and those in the evening targeted the social exchange processes during the reflection about the engagement in MVPA. Depending on the timing of the JITAIs, the expected effect might not manifest on the same day the couple received the dyadic JITAI but may be on a subsequent day they intended to engage in MVPA (Figures S2-S4 in Multimedia Appendix 4).

Each day, a randomization process either selected a JITAI or not, the latter serving as a control condition. Multimedia Appendix 4 provides a description of the dyadic JITAIs and the selection process. All interventions were stand-alone interventions and were triggered automatically without human involvement.

Measures

Intervention Variables

This study comprised several variables for the different intervention components. First, the study was categorized into the control phase (ie, phase without any dyadic interventions, serving as the reference category) and intervention phase (ie, phase with dyadic interventions). Second, planning was categorized into days without planned physical activities (coded as 0) and days for which the couples had planned physical activities during the planning intervention (coded as 1). Third, regarding the dyadic JITAIs, we included various dummy variables to describe whether a specified JITAI targeted the MVPA that day or did not target the day. We defined two variables for JITAIs targeting the actor and the partner: (1) the

dyadic JITAI for the actor compared days on which a dyadic JITAI targeted the actor's MVPA (coded as 1) with days without dyadic JITAI targeting the actor's MVPA (coded as 0), and (2) the dyadic JITAI for the partner compared days on which a dyadic JITAI targeted the partner's MVPA (coded as 1) with days without dyadic JITAI targeting the partner's MVPA (coded as 0). For the exploratory analyses, we created two variables representing the types of JITAI: (1) cross-over JITAI (coded as 1) versus no cross-over JITAI (coded as 0) targeting the MVPA on that day, and (2) joint JITAI (coded as 1) versus no joint JITAI (coded as 0) targeting the MVPA on that day. Joint JITAIs were coded as targeting both the actor's and partner's MVPA. Moreover, we created three variables representing the timing of the JITAI: (1) JITAI before the planning intervention (coded as 1) versus no JITAI before the planning intervention (coded as 0) targeting the MVPA that day, (2) JITAI before the activity (coded as 1) versus no JITAI before the activity (coded as 0) targeting the MVPA that day, and (3) JITAI in the evening (coded as 1) versus no JITAI in the evening (coded as 0) targeting the MVPA that day. Finally, we computed three lagged terms of the JITAI variables to describe how many occasions ago the JITAI targeted the MVPA: (1) JITAI targeting the MVPA 1 occasion before (coded as 1) compared with no JITAI targeting the MVPA one occasion before (coded as 0), (2) JITAI targeting the MVPA 2 occasions before (coded as 1) compared with no JITAI targeting the MVPA 2 occasions before (coded as 0), and (3) JITAI targeting the MVPA 3 occasions before (coded as 1) compared with no JITAI targeting the MVPA 3 occasions before (coded as 0).

Physical Activity

Physical activity was assessed using both device-based and self-reported methods. The device-based physical activity in minutes of MVPA per day was measured using wrist-worn accelerometers (ActiGraph CentrePoint Insight Watches [Ametris]). The acceleration is measured on 3 axes from which a single vector magnitude count is calculated. Values above 2690 counts per minute were classified as MVPA [44]. The device-based MVPA was filtered for wear time of at least 10 hours per day, according to Choi et al [45], and for awake time, according to the algorithm proposed by Tracy et al [46]. Overall, 11.3% (n=474) of the responses on device-based MVPA are missing after excluding days with low wear-compliance. The self-reported MVPA was measured by adding up the 2 questions capturing the MVPA they did alone ("How many minutes did you spend today alone doing moderate-to-vigorous physical activity?") and with their partner ("How many minutes did you spend today together with your partner doing moderate-to-vigorous physical activity?"). Overall, 5.6% (n=236) of the responses for self-reported MVPA are missing.

Covariates

We assessed the intervention group with 2 time-invariant dummy covariates (group A serving as the reference category) to address differences due to different sequences in the intervention and control phases. Furthermore, we assessed various time-variant covariates. We included the wear time of the wrist-worn accelerometer (minutes per day) because increased wear time increases the potential duration of MVPA that can be recorded. Moreover, we included the time (per 7 days) to address potential time trends [47]. Furthermore, we included dummy variables for the weekend (weekday serving as the reference category) to address differences in the behavior between weekdays and weekends, and skilled support intervention (before the intervention serving as the reference category) to address the potential effects of this intervention on the MVPA. Finally, we assessed the barriers and facilitating factors. Barriers and facilitating factors were assessed with the item "What has made your physical activity easier or more difficult today?" on a bipolar scale from −5 to 5, on which the participants responded to 8 influences that hindered or facilitated engagement in physical activity (eg, "hindering conditions for physical activity" to "beneficial conditions for physical activity"). The scores of the barriers and facilitating factors were summed up into 2 separate scores, indicating the total barriers and facilitating factors today. We included the barriers and facilitating factors to control the internal and external influences that may change the probability or duration of the engagement in MVPA and to test if the interventions effectively increased MVPA over and above these factors [48].

Data Analysis

To estimate the treatment effect of the intervention phase, we calculated multilevel models for intensive longitudinal data and indistinguishable dyads [49]. This approach can be used to analyze dyadic data by accounting for the nonindependence of observations within dyads. We included the intervention phase as the predictor variable and the device-based or self-reported MVPA as the outcome variables.

We used weighted and centered least-squares estimation for microrandomized trials to estimate the treatment effects of the planning interventions and the dyadic JITAIs [50]. This approach addresses the potential biases of time-varying treatment effects and provides consistent causal effects in microrandomized trials. We extended this approach to dyadic data by considering the dyads as the units of analysis. To be inclusive of same-gender couples in our study, we treated the partners as indistinguishable rather than distinguishable based on gender [49]. We included the planning or the JITAI targeting the actor and partner as predictor variables, respectively. We ran separate models with device-based and self-reported MVPA as outcome variables. In all models, complete case analyses were conducted. Thus, days with missing data were excluded from the data analyses. We included time as a covariate in all models [47]. Additionally, in analyses of device-based MVPA, we included the wear time of the wrist-worn accelerometer as an additional covariate.

Furthermore, we conducted sensitivity analyses to account for potential influences of covariates. In these sensitivity analyses, we additionally included variables indicating the intervention group, weekend, skilled support intervention, barriers, and facilitating factors as covariates. All covariates were grand mean-centered. All analyses were complete-case analyses.

We ran 3 separate models for the exploratory analyses. In these exploratory analyses, we included the dyadic JITAI type (ie, cross-over and joint), the JITAI timing (ie, before planning,

before the activity, and evening), and the lagged terms of the JITAI as dummy variables.

Software

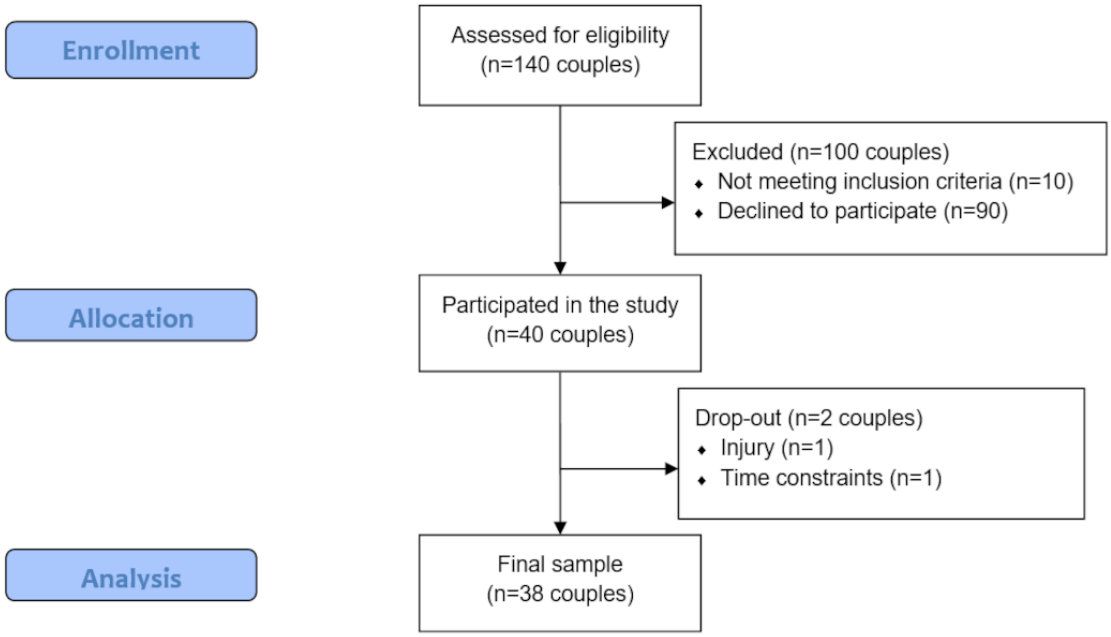
The analyses were conducted with R in RStudio [51,52]. We used the nlme package (v3.1-157) to estimate the effect of the intervention phase [53] and the xgeepack package (v1.3.9) to estimate the causal treatment effects of dyadic and collaborative planning and dyadic JITAIs [50,54,55]. All R codes are provided on OSF.

Results

Overview

Initially, 140 couples signed up for the study. However, 100 of them either chose not to participate or were excluded during screening. Thus, a total of 40 couples (80 participants) participated in the study, which accounted for an expected attrition rate of 20%. The recruitment stopped when the planned sample size was reached. Furthermore, 1 couple dropped out of the study due to an injury of 1 partner, and 1 couple discontinued the study because of time issues, leaving a total sample of 38 couples (37 mixed-gender and 1 same-gender couple). The CONSORT (Consolidated Standards of Reporting Trials) flow diagram is shown in Figure 1. The mean duration of MVPA was 116.64 (SD 35.40) minutes per day for device-based MVPA and 30.24 (SD 22.19) minutes per day for self-reported MVPA. There is a moderate to high correlation between the device-based and self-reported MVPA, with r_b=0.57 (P<.001) at the between-person and r_w=0.46 (P<.001) at the within-person level.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.



Tables 2 and 3 present the effects of the intervention phase on device-based and self-reported MVPA, expressed as the difference in minutes between days during the intervention phase compared with days during the control phase. In line with hypothesis 1, the results indicated that during the intervention phase, the couples had significantly higher device-based and self-reported MVPA than during the control phase. Furthermore, the intercepts and slopes showed significant variability, indicating heterogeneity in the mean levels of MVPA and the effects of the intervention phases on the MVPA between the couples.

Table 2. Effects of the intervention phase on device-based moderate-to-vigorous physical activity.

Effect[a]	Estimate, b (SE)	90% CI	*t* test (*df*) or *z* score	*P* value[b]
Fixed effects				
Intercept	106.50 (4.39)	99.29 to 113.72	24.28 (3665)[c]	<.001
Intervention phase[d]	5.88 (3.04)	0.88 to 10.89	1.93 (3665)[c]	.03
Time[e]	−1.05 (0.85)	−2.44 to 0.34	−1.24 (3665)[c]	.22
Wear time[f]	0.15 (0.01)	0.13 to 0.17	13.04 (3665)[c]	<.001
Random effects				
Intercept	24.30 (3.30)	19.48 to 30.33	7.37[g]	<.001
Intervention phase	8.59 (3.60)	4.51 to 16.36	2.39[g]	.02
Time	4.21 (0.83)	3.06 to 5.79	5.06[g]	<.001
Wear time	0.06 (0.01)	0.05 to 0.08	5.69[g]	<.001
Residuals	49.89 (0.65)	48.83 to 50.97	76.53[g]	<.001
Autocorrelation	0.27 (0.02)	0.24 to 0.30	15.67[g]	<.001

[a]Number of couples=38; number of days=55; number of cases in the analysis=3706.

[b]All *P* values are 2-tailed except those of the hypothesized effect of the intervention phase, where 1-tailed *P* values are used.

[c]*t* test.

[d]Intervention phase (coded as 1) included all days during the period when the couple received dyadic interventions and was compared with the control phase (coded as 0).

[e]Time was grand mean-centered per 7 days.

[f]Wear time of the accelerometers in minutes was grand mean-centered.

[g]*z* score.

Table 3. Effects of the intervention phase on self-reported moderate-to-vigorous physical activity.

Effects[a]	Estimate, b (SE)	90% CI	*t* test (*df*) or *z* score	*P* value[b]
Fixed effects				
Intercept	26.65 (3.17)	21.44 to 31.86	8.41 (3904)[c]	<.001
Intervention phase[d]	8.26 (3.88)	1.88 to 14.63	2.13 (3904)[c]	.02
Time[e]	−2.28 (0.95)	−3.84 to −0.71	−2.40 (3904)[c]	.02
Random effects				
Level-2				
Intercept	16.95 (2.50)	13.33 to 21.54	6.79[f]	<.001
Intervention phase	19.01 (3.50)	14.11 to 25.62	5.43[f]	<.001
Time	5.20 (0.79)	4.06 to 6.67	6.55[f]	<.001
Level-1				
Residuals	49.14 (0.57)	48.22 to 50.08	86.93[f]	<.001
Autocorrelation	0.03 (0.02)	0.00 to 0.06	1.74[f]	.08

[a]Number of couples=38; number of days=55; number of cases in the analysis=3994.

[b]All *P* values are 2-tailed except those of the hypothesized effect of the intervention phase, where 1-tailed *P* values are used.

[c]*t* test.

[d]Intervention phase (coded as 1) included all days during the period when the couple received dyadic interventions and was compared to the control phase (coded as 0).

[e]Time was grand mean-centered per 7 days.

[f]*z* score.

The average treatment effects of the planning interventions on device-based and self-reported MVPA in minutes are presented in Table 4. The results indicate that on days the participants planned to be physically active during the planning interventions, they indicated higher self-reported but not device-based MVPA. Thus, hypothesis 2 is only supported for self-reported but not device-based MVPA.

Table 4. Effects of the dyadic and collaborative planning interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a]	Device-based MVPA[b,c]			Self-reported MVPA[d]		
	Estimate, b (SE)	90% CI	*P* value[e]	Estimate, b (SE)	90% CI	*P* value[e]
Intercept	107.59 (3.79)	101.19 to 114.00	<.001	30.27 (3.51)	24.34 to 36.21	<.001
Planning[f]	6.31 (5.18)	−2.45 to 15.06	.12	14.25 (5.16)	5.53 to 22.98	.005
Time[g]	−1.59 (0.87)	−3.06 to −0.12	.08	−2.27 (0.85)	−3.71 to −0.83	.01
Wear time[h]	0.16 (0.01)	0.13 to 0.18	<.001	—[i]	—	—

[a]Number of couples=38; number of days=55.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[d]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[e]All *P* values are 2-tailed except those of the hypothesized effects of the planning interventions, where 1-tailed *P* values are used.

[f]Days with planned physical activities (coded as 1) were compared with days without any planned physical activities (coded as 0).

[g]Time was grand mean-centered per 7 days.

[h]Wear time of the accelerometers in minutes was grand mean-centered.

[i]Not applicable.

The average treatment effects of the dyadic JITAIs on device-based and self-reported MVPA in minutes are presented in Table 5. In line with hypothesis 3, the results indicated that both the dyadic JITAIs targeting the actor as well as the partner significantly increased the device-based and self-reported MVPA.

Table 5. Effects of the dyadic just-in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a]	Device-based MVPA[b,c]			Self-reported MVPA[d]		
	Estimate, b (SE)	90% CI	*P* value[e]	Estimate, b (SE)	90% CI	*P* value[e]
Intercept	107.62 (3.73)	101.31 to 113.93	<.001	30.17 (3.47)	24.30 to 36.04	<.001
Dyadic $JITAI_{Actor}$[f,g]	11.17 (3.18)	5.79 to 16.55	<.001	17.34 (3.65)	11.17 to 23.51	<.001
Dyadic $JITAI_{Partner}$[h]	7.23 (3.60)	1.14 to 13.33	.03	11.82 (4.10)	4.89 to 18.75	.003
Time[i]	−1.64 (0.79)	−2.97 to −0.30	.046	−2.17 (0.73)	−3.41 to −0.93	.006
Wear time[j]	0.16 (0.01)	0.13 to 0.18	<.001	—[k]	—	—

[a]Number of couples=38; number of days=55.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[d]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[e]All *P* values are 2-tailed except those of the effects of the hypothesized dyadic just-in-time adaptive interventions, where 1-tailed *P* values are used.

[f]JITAI: just-in-time adaptive intervention.

[g]Days on which a dyadic just-in-time adaptive intervention targeted the actor's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the actor's moderate-to-vigorous physical activity (coded as 0).

[h]Days on which a dyadic just-in-time adaptive intervention targeted the partner's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the partner's moderate-to-vigorous physical activity (coded as 0).

[i]Time was grand mean-centered per 7 days.

[j]Wear time of the accelerometers in minutes was grand mean-centered.

[k]Not applicable.

Sensitivity Analyses

Results of the sensitivity analyses, including the effect of the intervention phase, the average treatment effects of the planning, and the dyadic JITAIs, are reported in Tables S1-S3 in Multimedia Appendix 5. The pattern of the effects of the intervention phase on the device-based and self-reported MVPA remained the same after controlling for various covariates, suggesting the robustness of the results. However, the average treatment effect of the planning intervention on self-reported MVPA disappeared when controlling for covariates. Similarly, the average treatment effect of the dyadic JITAIs targeting the actor on the self-reported MVPA was no longer significant when adding the covariates. We further explored which covariates were responsible for these different results compared with the main analyses. Excluding the barriers and facilitating factors from the sensitivity analyses led to the same patterns as in the main analyses, suggesting that the differences were mainly driven by the barriers and facilitating factors (Table S4 in Multimedia Appendix 5). Moreover, the skilled support intervention resulted in a negative effect on the self-reported MVPA in some sensitivity analyses.

Exploratory Analyses

We conducted 3 exploratory analyses to gain insights into the effects of the type, timing, and temporal dynamics of the dyadic JITAIs (Tables 6-8). Regarding the type of dyadic JITAIs (ie, cross-over or joint), the results showed significant effects of cross-over JITAIs for both targeting the actor as well as the partner on device-based and self-reported MVPA. That means that on days when cross-over dyadic JITAIs were sent, couples were more physically active. Joint JITAIs did not increase device-based but increased self-reported MVPA, indicating that on days when joint dyadic JITAIs were sent, couples reported being more physically active. Furthermore, the effect sizes of the cross-over and joint JITAIs were comparable for both device-based and self-reported MVPA.

Table 6. Effects of the exploratory analyses of cross-over and joint just-in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a]	Device-based MVPA[b,c]			Self-reported MVPA[d]		
	Estimate, b (SE)	90% CI	*P* value[e]	Estimate, b (SE)	90% CI	*P* value[e]
Intercept	107.64 (3.75)	101.30 to 113.98	<.001	30.17 (3.49)	24.27 to 36.08	<.001
Cross-over $JITAI_{Actor}$[f,g]	12.31 (4.16)	5.26 to 19.34	.006	13.68 (4.24)	6.51 to 20.86	.003
Cross-over $JITAI_{Partner}$[h]	8.68 (4.24)	1.50 to 15.86	.049	11.91 (5.22)	3.09 to 20.74	.03
Joint JITAI[i]	8.87 (6.00)	−1.30 to 19.04	.15	19.20 (9.56)	3.03 to 35.69	.05
Time[j]	−1.59 (0.78)	−2.91 to −0.28	.049	−2.06 (0.78)	−3.37 to −0.75	.01
Wear time[k]	0.16 (0.01)	0.13 to 0.18	<.001	—[l]	—	—

[a]Number of couples=38; number of days=55.

[b]MVPA: moderate-to-vigorous physical activity.

[c]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[d]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[e]All *P* values are 2-tailed.

[f]JITAI: just-in-time adaptive intervention.

[g]Days on which a cross-over JITAI targeted the actor's MVPA (coded 1) were compared to days without cross-over JITAIs targeting the actors's MVPA (coded 0).

[h]Days on which a cross-over JITAI targeted the partner's MVPA (coded 1) were compared to days without cross-over JITAIs aiming at the partner's MVPA (coded 0).

[i]Days on which a joint JITAI targeted both's MVPA (coded 1) were compared to days without joint JITAIs (coded 0).

[j]Time was grand-mean-centered per 7 days.

[k]Wear time of the accelerometers in minutes was grand-mean-centered.

[l]Not applicable.

Table 7. Effects of the exploratory analyses of the timing of the dyadic just in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a,b,c]	Device-based MVPA[d,e]			Self-reported MVPA[f]		
	Estimate, b (SE)	90% CI	*P* value[g]	Estimate, b (SE)	90% CI	*P* value[g]
Intercept	107.65 (3.72)	101.34 to 113.97	<.001	30.16 (3.45)	24.30 to 36.02	<.001
Dyadic JITAI[h] planning$_{Actor}$	16.59 (4.74)	8.54 to 24.64	.002	17.92 (6.15)	7.48 to 28.36	.007
Dyadic JITAI activity$_{Actor}$	−6.82 (5.37)	−15.94 to 2.30	.21	10.87 (6.39)	0.03 to 21.71	.10
Dyadic JITAI evening$_{Actor}$	4.17 (3.29)	−1.42 to 9.77	.22	−1.69 (6.11)	−12.06 to 8.68	.74
Dyadic JITAI planning$_{Partner}$	6.57 (5.11)	−2.11 to 15.25	.21	13.97 (6.18)	3.47 to 24.46	.03
Dyadic JITAI activity$_{Partner}$	8.07 (5.65)	−1.53 to 17.67	.16	9.55 (2.99)	4.48 to 14.62	.003
Dyadic JITAI evening$_{Partner}$	3.54 (3.82)	−2.95 to 10.03	.36	5.41 (3.43)	−0.41 to 11.24	.12
Time[i]	−1.66 (0.77)	−2.98 to −0.34	.04	−2.15 (0.69)	−3.33 to −0.98	.004
Wear time[j]	0.16 (0.01)	0.13 to 0.18	<.001	—[k]	—	—

[a]Number of couples=38; number of days=55.

[b]The “Actor” and “Partner” subscripts indicate who was targeted by the dyadic just-in-time adaptive intervention. For just-in-time adaptive interventions targeting the actor, days on which a dyadic just-in-time adaptive intervention targeted the actor’s moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the actor’s moderate-to-vigorous physical activity (coded as 0). For just-in-time adaptive interventions targeting the partner, days on which a dyadic just-in-time adaptive intervention targeted the partner’s moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the partner’s moderate-to-vigorous physical activity (reference category, coded as 0).

[c]Dyadic just-in-time adaptive interventions may be triggered before the planning intervention (labeled planning), before a planned activity (labeled activity), and in the evening (labeled evening).

[d]MVPA: moderate-to-vigorous physical activity.

[e]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[f]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[g]All *P* values are 2-tailed.

[h]JITAI: just-in-time adaptive intervention.

[i]Time was grand-mean-centered per 7 days.

[j]Wear time of the accelerometers in minutes was grand mean-centered.

[k]Not applicable.

Table 8. Effects of the exploratory analyses of the lagged effects of the dyadic just-in-time adaptive interventions on device-based and self-reported moderate-to-vigorous physical activity.

Parameter[a,b,c]	Device-based MVPA[d,e]			Self-reported MVPA[f]		
	Estimate, b (SE)	90% CI	*P* value[g]	Estimate, b (SE)	90% CI	*P* value[g]
Intercept	107.63 (3.73)	101.28 to 113.97	<.001	30.05 (3.51)	24.08 to 36.02	<.001
Dyadic $JITAI_{Actor}$[h]	7.79 (3.12)	2.47 to 13.10	.02	9.23 (3.44)	3.38 to 15.07	.01
Dyadic $JITAI_{Actor\ lag1}$	5.69 (3.28)	0.10 to 11.27	.09	10.71 (3.41)	4.90 to 16.51	.004
Dyadic $JITAI_{Actor\ lag2}$	0.22 (2.55)	−4.13 to 4.56	.93	6.38 (2.65)	1.87 to 10.89	.02
Dyadic $JITAI_{Actor\ lag3}$	1.13 (4.27)	−6.14 to 8.40	.79	5.25 (5.16)	−3.52 to 14.03	.31
Dyadic $JITAI_{Partner}$	4.10 (4.04)	−2.79 to 10.99	.32	7.66 (4.27)	0.39 to 14.92	.08
Dyadic $JITAI_{Partner\ lag1}$	3.22 (2.87)	−1.68 to 8.11	.27	4.35 (2.68)	−0.20 to 8.91	.12
Dyadic $JITAI_{Partner\ lag2}$	1.61 (3.18)	−3.80 to 7.03	.62	−0.96 (2.76)	−5.65 to 3.73	.73
Dyadic $JITAI_{Partner\ lag3}$	0.44 (4.10)	−6.54 to 7.42	.92	−2.33 (4.35)	−9.73 to 5.07	.60
Time[i]	−1.82 (0.83)	−3.23 to −0.41	.04	−2.56 (0.73)	−3.81 to −1.31	.002
Wear time[j]	0.16 (0.01)	0.13 to 0.18	<.001	—[k]	—	—

[a]Number of couples=38; number of days=55.

[b]The "Actor" and "Partner" subscripts indicate who was targeted by the dyadic just-in-time adaptive intervention. For just-in-time adaptive interventions targeting the actor, days on which a dyadic just-in-time adaptive intervention targeted the actor's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the actor's moderate-to-vigorous physical activity (coded as 0). For just-in-time adaptive interventions targeting the partner, days on which a dyadic just-in-time adaptive intervention targeted the partner's moderate-to-vigorous physical activity (coded as 1) were compared with days without dyadic just-in-time adaptive interventions targeting the partner's moderate-to-vigorous physical activity (coded as 0).

[c]The 3 lag parameters indicate how many occasions ago the just-in-time adaptive interventions targeted the moderate-to-vigorous physical activity.

[d]MVPA: moderate-to-vigorous physical activity.

[e]Number of cases in the analysis for device-based moderate-to-vigorous physical activity=3706.

[f]Number of cases in the analysis for self-reported moderate-to-vigorous physical activity=3944.

[g]All *P* values are 2-tailed.

[h]JITAI: just-in-time adaptive intervention.

[i]Time was grand mean-centered per 7 days.

[j]Wear time of the accelerometers in minutes was grand mean-centered.

[k]Not applicable.

The dyadic JITAIs varied in effectiveness depending on the time they were sent. The dyadic JITAIs sent before the planning intervention targeting the actor increased the device-based and self-reported MVPA. The dyadic JITAIs sent before the planning intervention targeting the partner increased the self-reported but not the device-based MVPA. Furthermore, the dyadic JITAIs sent before the planned activity targeting the actor did not increase device-based nor self-reported MVPA. The dyadic JITAIs sent before the planned activity targeting the partner increased self-reported, but not the device-based MVPA. Finally, neither the dyadic JITAIs sent in the evening targeting the actor nor the partner had any significant effect on the self-reported or device-based MVPA.

Regarding the temporal dynamics of the JITAIs, there was evidence for lagged effects of the dyadic JITAIs targeting the actor on self-reported and, to a lesser extent, on device-based MVPA. This finding indicates that the dyadic JITAIs targeting the actor not only affected the MVPA at the time they were initially aimed at, but their effects lasted and influenced subsequent self-reported MVPA. There was no evidence of lagged effects of dyadic JITAIs targeting the partner on device-based or self-reported MVPA.

Discussion

Principal Findings

This study demonstrated the promising effects of smartphone-based dyadic interventions in promoting MVPA in romantic partners. Participants reported higher device-based and self-reported MVPA during the intervention phase than during the control phase. Furthermore, planning physical activities increased the participants' self-reported but not device-based MVPA. Dyadic JITAIs targeting the actor as well as the partner increased the device-based and self-reported MVPA. However, sensitivity analyses indicated that some of these effects were not robust, suggesting that contextual factors play a relevant role in the effectiveness of dyadic JITAIs. Exploratory analyses showed that cross-over JITAIs promoted device-based and self-reported MVPA, and joint JITAIs promoted self-reported MVPA. Moreover, dyadic JITAIs aiming

at different times (ie, before a planning intervention, before a planned activity, and in the evening) were differently effective in increasing MVPA. Finally, there was evidence for lagged effects of the dyadic JITAIs targeting the actor for device-based and self-reported MVPA but not for lagged effects of dyadic JITAIs targeting the partner.

Effects of the Dyadic Interventions

Consistent with past research, this study found promising effects of dyadic interventions to increase physical activity. Overall, the complex intervention, including different dyadic interventions, was effective in increasing both device-based and self-reported MVPA (hypothesis 1). However, there was heterogeneity in the effects of the intervention phase across couples. This heterogeneity may reflect differences in couple dynamics or insufficient knowledge of moderating factors interacting with the intervention or control phase [48]. A better understanding of these factors may help improve the effectiveness of dyadic interventions [39,56]. Future research may investigate intrapersonal characteristics, relationship dynamics, and the broader social context to provide valuable insights into factors relevant to intervention development and implementation [56].

This study found significant effects of the planning intervention on the self-reported but not device-based MVPA (hypothesis 2). These results are comparable with previous studies that found mixed effects of dyadic and collaborative planning [28,31,32,35]. In contrast to previous studies, this study used a within-couple design and compared days with planned activities to days without planned activities instead of comparing the effects of various planning interventions (eg, dyadic, collaborative, and individual) on subsequent physical activity. Additionally, the participants could freely choose and vary if they wanted to plan dyadically or collaboratively. This approach aimed to enable participants to plan according to their needs and preferences to compensate for potential challenges associated with each form of planning. However, these attempts to better tailor the planning interventions to couples' needs and to use more contingent outcomes [57] were not sufficient to enhance the planning intervention's effectiveness, at least with regard to the device-based measure of MVPA. As discussed previously [58], it seems necessary to identify under what circumstances and for whom these forms of planning unfold the most beneficial effects.

To our knowledge, our study was the first to implement dyadic JITAIs to promote MVPA in romantic couples. In line with previous studies on the effectiveness of JITAIs on health behaviors [39], this study showed promising effects of dyadic JITAIs in promoting MVPA (hypothesis 3a). Furthermore, not only the dyadic JITAIs targeting the actor but also those targeting the partner increased the MVPA (hypothesis 3b). This finding may be explained by both the dyadic nature of the JITAIs as well as the interdependence between close partners in a romantic relationship [6]. One partner's engagement may motivate the other partner to engage in physical activity as well [59]. Furthermore, helping the partner engage in physical activity may have positive outcomes for the provider [43]. For example, providing social support or control to the partner may reiterate the importance of physical activity for the provider [60], increase self-esteem [61], and contribute to feeling more energized [62], promoting engagement in physical activity. Thus, deciding who to target with dyadic interventions appears to be an essential design choice when developing and implementing dyadic interventions.

The effect of the planning and the effect of the dyadic JITAIs targeting the actor on self-reported MVPA disappeared when controlling for additional covariates. Exploratory analyses suggested that these discrepancies were driven by the reported barriers and facilitating factors. Potentially, the barriers and facilitating factors may act as mediators in explaining the JITAIs' effects on MVPA. The planning and JITAIs prompted processes to help engage in MVPA, which may have removed barriers or facilitated engagement in MVPA. Furthermore, it is also possible that the causality goes in the other direction, in that engagement and nonengagement in MVPA influenced the perception of barriers and facilitating factors.

We found an unexpected negative effect of the skilled support intervention on self-reported MVPA in some of the analyses. However, it is important to note that this part of the intervention was not randomized.

This study found some differences in the effects of the dyadic interventions on device-based and self-reported MVPA. Interestingly, in this study, the device-based measure of MVPA was much higher than the self-reported measure. These 2 measures represent related but distinct indicators of MVPA [63]. There are several explanations for these differences. First, self-reported and device-based MVPA may capture different aspects of physical activity [64]. For example, self-reported physical activity may refer more closely to the activities perceived as physical activity by the participants, whereas device-based physical activity may also include other forms of movement, such as naturally-occurring MVPA. Furthermore, it may be possible that engaging in planned physical activities (captured by both the device-based and self-reported MVPA) leads to compensatory effects in that the engagement in naturally occurring MVPA (captured only by device-based MVPA) decreases [65], leading to some discrepancies between the measures. Second, there may be response biases in the self-reported but not device-based MVPA measure [64]. Participants may indicate higher MVPA if they received an intervention because they expect it to be effective. Finally, the device-based MVPA may have misclassified some motions as MVPA, which would also explain why the device-based MVPA was substantially higher than the self-reported MVPA [65].

Exploratory Analyses

Results from the exploratory analyses showed that cross-over JITAIs increased device-based and self-reported MVPA, and joint JITAIs increased self-reported MVPA while having similar effect sizes. Given the exploratory nature of our analyses, results need to be replicated in confirmatory designs and analyses in the future. In addition to examining the main effects of these different prototypes of dyadic interventions [10] and despite their comparable effect sizes found in our study, these 2 types of interventions may target different mechanisms of action [10,14]. For example, joint interventions, such as discussions

and collaborative planning, may trigger interactions in which both partners are equally involved. In contrast, cross-over interventions, such as prompting to provide social support and to exert social control, may trigger one-sided interactions where one partner is more involved than the other. Future research is needed to better understand the potential different effects of cross-over and joint interventions and how they may vary under specific circumstances.

The second set of exploratory analyses showed that dyadic JITAIs with different timings varied in effectiveness. Specifically, there was stronger evidence for the effectiveness of the dyadic JITAIs sent before the planning intervention and before the planned activity than those sent in the evening. These dyadic JITAIs targeted different aspects of the behavior change process. The dyadic JITAIs before the planning interventions aimed at increasing the quality of the dyadic and collaborative plans, a factor central to the effectiveness of planning [58]. The dyadic JITAIs before the planned activity aimed at helping to engage in this activity. Sending dyadic JITAIs at this specific time may be effective because they directly enhance commitment to the planned activities. The dyadic JITAIs in the evening aimed at the partners to reflect on their goal progress, discussing things that went well or poorly regarding physical activity, or prompting support or positive control in the future. A reason for the lack of effectiveness may be that these dyadic JITAIs were not matched to a particular state of vulnerability or opportunity since they did not target immediate preparation or engagement in MVPA directly. Furthermore, the participants may not have had the opportunity to engage in the prompted task because the interventions were sent too late in the day. Thus, future studies should also examine intervention fidelity and intervention engagement, which may vary with different timings of the interventions.

The final exploratory analyses showed that dyadic JITAIs may have effects over and above their targeted time point. Implementing dyadic JITAIs to improve the social exchange processes may translate to improved interaction patterns that sustainably facilitate engagement in health behaviors [11]. Since many health behaviors, such as physical activity, require consistent engagement over extended periods of time, this would be promising for promoting longer-term behavior change [66]. Specifically, this study found evidence for long-term effects of dyadic JITAIs targeting the actor but not for those targeting the partner. These results suggest that different mechanisms may explain these effects [43]. For example, receiving social support and control may establish a subjective norm or change attitudes regarding physical activity [67], which translates into regular engagement in physical activity. In contrast, providing social support and control may promote MVPA by increasing positive effects [43] and energy levels [61], which may entail more transient effects. Future research is needed to investigate these underlying mechanisms in more detail.

Past research has illustrated a need to better understand the boundary conditions moderating the effectiveness of (dyadic) JITAIs [39]. The exploratory analyses of this study contribute to this call by illustrating preliminary insights into the importance of the type and timing of the dyadic JITAIs for their effectiveness. Additionally, it showed the potential long-term effects of dyadic JITAIs, providing new insights and opportunities for health behavior promotion. Given the exploratory nature, however, the next steps are to replicate findings with confirmatory analyses. Moreover, in future studies with more participants, it might also be worthwhile to examine interactions between the different factors examined here. For example, it might well be that joint JITAIs work better when being sent in the evening to prompt joint reflection of past interactions, compared with cross-over JITAIs that might work better before a planning intervention. It is also possible that joint JITAIs might have longer-lasting effects, as the positive dyadic dynamics in supportive interactions promoted by these kinds of dyadic JITAIs might add to the maintenance of joint behavior change attempts.

Strengths and Limitations

This study has various strengths. First, the study investigated health behavior change from a dyadic rather than individual perspective, allowing for a more complete understanding of health behavior change processes in romantic relationships. Second, the microrandomized trial design allowed us to experimentally investigate the dyadic health behavior change as a process over time and to examine daily dynamics [47]. This also allowed us to examine the causal effects of the interventions on the outcomes [50]. Third, we were able to investigate shorter-term as well as lagged effects of the dyadic JITAIs. Finally, including device-based and self-reported MVPA measurements offered a more comprehensive understanding of the effects of dyadic interventions on MVPA [63], as well as addressed potential methodological challenges inherent to the measurement methods.

Despite these strengths, this study also has its limitations. As it was a pilot study, the sample size of 38 romantic couples was relatively small. This relatively small sample size may limit the statistical power, which may be especially the case for the exploratory analyses, where there have been fewer instances of interventions. However, the relatively large number of days examined per couple (ie, 55 days) partly compensates for this small sample size [47]. The limited number of couples also restricted the possibility of conducting more detailed analyses exploring the effectiveness of interventions targeting specific social exchange processes (eg, emotional support and positive social control) and potential mechanisms underlying the intervention effects or interaction effects as outlined above. Furthermore, there were different types of dyadic interventions (ie, skilled support intervention, planning interventions, and JITAIs) implemented in the study. These interventions were designed to build upon and facilitate each other. For example, the dyadic JITAIs were built on the knowledge of skilled support and were triggered in reference to the planned activities. However, this also implies that it remains unclear how exactly and to what extent these different dyadic interventions influenced each other (eg, if the dyadic JITAIs would have been effective without an initial skilled support intervention). Finally, the study targeted inactive romantic couples who expressed a willingness to increase their physical activity. Thus, the study sample was self-selected and therefore may not fully represent the general internet population.

Conclusion

This study extended the common approach of promoting health behavior change of individuals by including partners in the behavior change process. This allowed us to implement DBCTs that address social interactions between the partners. Overall, this study demonstrated the effectiveness of dyadic interventions, including planning interventions and dyadic JITAIs. Furthermore, the exploratory analyses provided first hints to the assumptions that conditions, such as the type, timing, and target, moderate the effectiveness of dyadic JITAIs, providing new insights into conceptual and design elements of dyadic JITAIs. Future research is needed to get a more comprehensive understanding of the effectiveness of dyadic interventions. Specifically, the moderating variables influencing the effectiveness of dyadic interventions and the mechanisms explaining dyadic interventions should be examined.

Acknowledgments

No generative artificial intelligence has been used.

Funding

This study, as well as PSH and JMA, was funded by the Swiss National Science Foundation (grant 10001C_197471). Furthermore, JMA was funded for part of this work by the Medical Research Council (MR/X003264/1 and MC_UU_00022/3) and Chief Scientist Office (SPHSU18).

Data Availability

Data are available upon reasonable request.

Authors' Contributions

Conceptualization: PSH, US
Formal analysis: PSH
Funding acquisition: RT, US
Methodology: PSH
Project administration: RT, US
Resources: RT, US
Supervision: US
Writing – original draft: PSH
Writing – review and editing: RT, JMA, PK, US

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 324 KB - mhealth_v14i1e67136_app1.pdf]

Multimedia Appendix 2
Informed Consent.
[DOCX File , 17 KB - mhealth_v14i1e67136_app2.docx]

Multimedia Appendix 3
Screenshots Study App.
[PPTX File , 267 KB - mhealth_v14i1e67136_app3.pptx]

Multimedia Appendix 4
Dyadic Just-in-time Adaptive Interventions.
[DOCX File , 336 KB - mhealth_v14i1e67136_app4.docx]

Multimedia Appendix 5
Additional Results.
[DOCX File , 43 KB - mhealth_v14i1e67136_app5.docx]

References

1. Physical activity. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/physical-activity [accessed 2025-12-16]
2. Feil K, Fritsch J, Rhodes RE. The intention-behaviour gap in physical activity: a systematic review and meta-analysis of the action control framework. Br J Sports Med 2023;57(19):1265-1271. [doi: 10.1136/bjsports-2022-106640] [Medline: 37460164]
3. Körperliche Aktivität. Bundesamt für Statistik. 2022. URL: https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/determinanten/koerperliche-aktivitaet.html [accessed 2025-12-16]
4. Gourlan M, Bernard P, Bortolon C, Romain AJ, Lareyre O, Carayol M, et al. Efficacy of theory-based interventions to promote physical activity. A meta-analysis of randomised controlled trials. Health Psychol Rev 2016;10(1):50-66. [doi: 10.1080/17437199.2014.981777] [Medline: 25402606]
5. Scholz U, Berli C, Lüscher J, Knoll N. Dyadic behavior change interventions. In: Hagger MS, Cameron LD, Hamilton K, Hankonen N, Lintunen T, editors. Handb Behav Change. England: Cambridge University Press; 2020:632-648.
6. Huelsnitz CO, Jones RE, Simpson JA, Joyal-Desmarais K, Standen EC, Auster-Gussman LA, et al. The dyadic health influence model. Pers Soc Psychol Rev 2022;26(1):3-34. [doi: 10.1177/10888683211054897] [Medline: 34873983]
7. Scholz U, Stadler G, Berli C, Lüscher J, Knoll N. How do people experience and respond to social control from their partner? Three daily diary studies. Front Psychol 2020;11:613546 [FREE Full text] [doi: 10.3389/fpsyg.2020.613546] [Medline: 33519637]
8. Cobb LK, Godino JG, Selvin E, Kucharska-Newton A, Coresh J, Koton S. Spousal influence on physical activity in middle-aged and older adults: The ARIC study. Am J Epidemiol 2016;183(5):444-451 [FREE Full text] [doi: 10.1093/aje/kwv104] [Medline: 26337074]
9. Carr RM, Prestwich A, Kwasnicka D, Thøgersen-Ntoumani C, Gucciardi DF, Quested E, et al. Dyadic interventions to promote physical activity and reduce sedentary behaviour: systematic review and meta-analysis. Health Psychol Rev 2019;13(1):91-109 [FREE Full text] [doi: 10.1080/17437199.2018.1532312] [Medline: 30284501]
10. Di Maio S, Villinger K, Knoll N, Scholz U, Stadler G, Gawrilow C, et al. Compendium of dyadic intervention techniques (DITs) to change health behaviours: a systematic review. Health Psychol Rev 2024;18(3):538-573 [FREE Full text] [doi: 10.1080/17437199.2024.2307534] [Medline: 38437798]
11. Richards EA, Franks MM, McDonough MH, Porter K. 'Let's move:' a systematic review of spouse-involved interventions to promote physical activity. International Journal of Health Promotion and Education 2017;56(1):51-67. [doi: 10.1080/14635240.2017.1415160]
12. Gellert P, Ziegelmann JP, Warner LM, Schwarzer R. Physical activity intervention in older adults: does a participating partner make a difference? Eur J Ageing 2011;8(3):211 [FREE Full text] [doi: 10.1007/s10433-011-0193-5] [Medline: 28798651]
13. Ramchand R, Ahluwalia SC, Xenakis L, Apaydin E, Raaen L, Grimm G. A systematic review of peer-supported interventions for health promotion and disease prevention. Prev Med 2017;101:156-170. [doi: 10.1016/j.ypmed.2017.06.008] [Medline: 28601621]
14. Rook KS, August KJ, Sorkin D. Social network functions and health. In: Handb Stress Sci Biol Psychol. Princeton, New Jersey: Health Springer Publishing Company; 2011:123-135.
15. Radtke T, Luszczynska A, Schenkel K, Biddle S, Scholz U. A cluster randomized controlled trial comparing the effectiveness of an individual planning intervention with collaborative planning in adolescent friendship dyads to enhance physical activity (TWOgether). BMC Public Health 2018;18(1):911 [FREE Full text] [doi: 10.1186/s12889-018-5818-6] [Medline: 30041603]
16. Cohen S. Social relationships and health. Am Psychol 2004;59(8):676-684. [doi: 10.1037/0003-066X.59.8.676] [Medline: 15554821]
17. Feeney BC, Collins NL. A new look at social support: a theoretical perspective on thriving through relationships. Pers Soc Psychol Rev 2015;19(2):113-147 [FREE Full text] [doi: 10.1177/1088868314544222] [Medline: 25125368]
18. Schwarzer R, Knoll N. Social support. In: Kaptein A, Weinman J, editors. Health Psychol 2nd ed Great Britain. Oxford: Blackwell; 2010:283-293.
19. Kouvonen A, De Vogli R, Stafford M, Shipley MJ, Marmot MG, Cox T, et al. Social support and the likelihood of maintaining and improving levels of physical activity: the whitehall II study. Eur J Public Health 2012;22(4):514-518 [FREE Full text] [doi: 10.1093/eurpub/ckr091] [Medline: 21750013]
20. Rackow P, Berli C, Lüscher J, Luszczynska A, Scholz U. Emotional or instrumental support? Distinct effects on vigorous exercise and affect. Psychology of Sport and Exercise 2017;33:66-74. [doi: 10.1016/j.psychsport.2017.07.011]
21. Scarapicchia TMF, Amireault S, Faulkner G, Sabiston CM. Social support and physical activity participation among healthy adults: a systematic review of prospective studies. International Review of Sport and Exercise Psychology 2016;10(1):50-83. [doi: 10.1080/1750984x.2016.1183222]
22. Craddock E, vanDellen MR, Novak SA, Ranby KW. Influence in relationships: A meta-analysis on health-related social control. Basic and Applied Social Psychology 2015;37(2):118-130. [doi: 10.1080/01973533.2015.1011271]

23. Lewis MA, Butterfield RM. Antecedents and reactions to health-related social control. Pers Soc Psychol Bull 2005;31(3):416-427. [doi: 10.1177/0146167204271600] [Medline: 15657456]
24. Rafaeli E, Gleason MEJ. Skilled support within intimate relationships. J of Family Theo & Revie 2009;1(1):20-37. [doi: 10.1111/j.1756-2589.2009.00003.x]
25. Shilts MK, Horowitz M, Townsend MS. Goal setting as a strategy for dietary and physical activity behavior change: a review of the literature. Am J Health Promot 2004;19(2):81-93. [doi: 10.4278/0890-1171-19.2.81] [Medline: 15559708]
26. Sniehotta F, Schwarzer R, Scholz U, Schüz B. Action planning and coping planning for long-term lifestyle change: theory and assessment. Eur. J. Soc. Psychol 2005;35(4):565-576 [FREE Full text] [doi: 10.1002/ejsp.258]
27. Wiedemann AU, Lippke S, Reuter T, Ziegelmann JP, Schwarzer R. How planning facilitates behaviour change: Additive and interactive effects of a randomized controlled trial. Eur. J. Soc. Psychol 2011;41(1):42-51. [doi: 10.1002/ejsp.724]
28. Szczuka Z, Kulis E, Boberska M, Banik A, Kruk M, Keller J, et al. Can individual, dyadic, or collaborative planning reduce sedentary behavior? A randomized controlled trial. Soc Sci Med 2021;287:114336 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114336] [Medline: 34482277]
29. Burkert S, Scholz U, Gralla O, Roigas J, Knoll N. Dyadic planning of health-behavior change after prostatectomy: a randomized-controlled planning intervention. Soc Sci Med 2011;73(5):783-792. [doi: 10.1016/j.socscimed.2011.06.016] [Medline: 21807446]
30. Prestwich A, Conner M, Lawton R, Bailey W, Litman J, Molyneaux V. Individual and collaborative implementation intentions and the promotion of breast self-examination. Psychology & Health 2005;20(6):743-760. [doi: 10.1080/14768320500183335]
31. Katzenelenbogen O, Knoll N, Stadler G, Bar-Kalifa E. The role of individual and dyadic planning in couples' daily goal pursuits. Pers Soc Psychol Bull 2022;48(2):239-253. [doi: 10.1177/0146167221997630] [Medline: 33783241]
32. Keller J, Fleig L, Hohl DH, Wiedemann AU, Burkert S, Luszczynska A, et al. Which characteristics of planning matter? Individual and dyadic physical activity plans and their effects on plan enactment. Soc Sci Med 2017;189:53-62. [doi: 10.1016/j.socscimed.2017.07.025] [Medline: 28783502]
33. Keller J, Hohl DH, Hosoya G, Heuse S, Scholz U, Luszczynska A, et al. Long-term effects of a dyadic planning intervention with couples motivated to increase physical activity. Psychology of Sport and Exercise 2020;49:101710. [doi: 10.1016/j.psychsport.2020.101710]
34. Knoll N, Hohl DH, Keller J, Schuez N, Luszczynska A, Burkert S. Effects of dyadic planning on physical activity in couples: A randomized controlled trial. Health Psychol 2017;36(1):8-20. [doi: 10.1037/hea0000423] [Medline: 27642760]
35. Kulis E, Szczuka Z, Keller J, Banik A, Boberska M, Kruk M, et al. Collaborative, dyadic, and individual planning and physical activity: A dyadic randomized controlled trial. Health Psychol 2022;41(2):134-144. [doi: 10.1037/hea0001124] [Medline: 34968130]
36. Prestwich A, Conner MT, Lawton RJ, Ward JK, Ayres K, McEachan RRC. Randomized controlled trial of collaborative implementation intentions targeting working adults' physical activity. Health Psychol 2012;31(4):486-495. [doi: 10.1037/a0027672] [Medline: 22468716]
37. 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;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
38. 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;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
39. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): A meta-analytical review. Health Commun 2020;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
40. OSF. URL: https://osf.io/d98yp/ [accessed 2026-01-15]
41. Time and Ties: Intervention Effects. OSF. URL: https://osf.io/q3vmu/ [accessed 2026-01-15]
42. Scholz U, Berli C. A dyadic action control trial in overweight and obese couples (DYACTIC). BMC Public Health 2014;14:1321 [FREE Full text] [doi: 10.1186/1471-2458-14-1321] [Medline: 25540972]
43. Berli C, Schwaninger P, Scholz U. "We Feel Good": daily support provision, health behavior, and well-being in romantic couples. Front Psychol 2020;11:622492 [FREE Full text] [doi: 10.3389/fpsyg.2020.622492] [Medline: 33536986]
44. Sasaki J, John D, Freedson P. Validation and comparison of ActiGraph activity monitors. J Sci Med Sport 2011;14(5):411-416. [doi: 10.1016/j.jsams.2011.04.003] [Medline: 21616714]
45. Choi L, Liu Z, Matthews C, Buchowski M. Validation of accelerometer wear and nonwear time classification algorithm. Med Sci Sports Exerc 2011;43(2):357-364 [FREE Full text] [doi: 10.1249/MSS.0b013e3181ed61a3] [Medline: 20581716]
46. Tracy JD, Acra S, Chen KY, Buchowski MS. Identifying bedrest using 24-h waist or wrist accelerometry in adults. PLoS One 2018;13(3):e0194461 [FREE Full text] [doi: 10.1371/journal.pone.0194461] [Medline: 29570740]
47. Bolger N, Laurenceau JP. Intensive Longitudinal Methods: an Introduction to Diary and Experience Sampling Research. New York: Guilford Press; 2013.
48. Rothman AJ, Sheeran P. The operating conditions framework: Integrating mechanisms and moderators in health behavior interventions. Health Psychol 2021;40(12):845-857. [doi: 10.1037/hea0001026] [Medline: 32914997]

49. Ledermann T, Kenny DA. Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. J Fam Psychol 2017;31(4):442-452. [doi: 10.1037/fam0000290] [Medline: 28165269]
50. Boruvka A, Almirall D, Witkiewitz K, Murphy SA. Assessing time-varying causal effect moderation in mobile health. J Am Stat Assoc 2018;113(523):1112-1121 [FREE Full text] [doi: 10.1080/01621459.2017.1305274] [Medline: 30467446]
51. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2020. URL: https://www.R-project.org/ [accessed 2025-12-16]
52. Posit Team. RStudio: integrated development environment for R. Posit Software, PBC. 2023. URL: http://www.posit.co/ [accessed 2025-12-16]
53. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: Linear and Nonlinear Mixed Effects Models. 2022. URL: https://CRAN.R-project.org/package=nlme [accessed 2025-12-16]
54. Qian T, Walton AE, Collins LM, Klasnja P, Lanza ST, Nahum-Shani I, et al. The microrandomized trial for developing digital interventions: experimental design and data analysis considerations. Psychol Methods 2022;27(5):874-894 [FREE Full text] [doi: 10.1037/met0000283] [Medline: 35025583]
55. Halekoh U, Højsgaard S, Yan J. The R package geepack for generalized estimating equations. J Stat Softw 2006;15(2). [doi: 10.18637/jss.v015.i02]
56. Rothman AJ, Simpson JA, Huelsnitz CO, Jones RE, Scholz U. Integrating intrapersonal and interpersonal processes: a key step in advancing the science of behavior change. Health Psychol Rev 2020;14(1):182-187. [doi: 10.1080/17437199.2020.1719183] [Medline: 31959071]
57. Sniehotta FF. Towards a theory of intentional behaviour change: plans, planning, and self-regulation. Br J Health Psychol 2009;14(Pt 2):261-273. [doi: 10.1348/135910708X389042] [Medline: 19102817]
58. Hagger MS, Luszczynska A. Implementation intention and action planning interventions in health contexts: state of the research and proposals for the way forward. Appl Psychol Health Well Being 2014;6(1):1-47. [doi: 10.1111/aphw.12017] [Medline: 24591064]
59. Jackson SE, Steptoe A, Wardle J. The influence of partner's behavior on health behavior change: the English longitudinal study of ageing. JAMA Intern Med 2015;175(3):385-392. [doi: 10.1001/jamainternmed.2014.7554] [Medline: 25599511]
60. Hoffman RK, Jobe MC, Dodge T. A brief self-persuasion intervention to strengthen health-promoting dietary intentions through autonomous motivation. Appetite 2023;180:106371. [doi: 10.1016/j.appet.2022.106371] [Medline: 36402411]
61. Nelson SK, Layous K, Cole SW, Lyubomirsky S. Do unto others or treat yourself? The effects of prosocial and self-focused behavior on psychological flourishing. Emotion 2016;16(6):850-861 [FREE Full text] [doi: 10.1037/emo0000178] [Medline: 27100366]
62. Liu Y, Kornfield R, Shaw BR, Shah DV, McTavish F, Gustafson DH. Giving and receiving social support in online substance use disorder forums: how self-efficacy moderates effects on relapse. Patient Educ Couns 2020;103(6):1125-1133 [FREE Full text] [doi: 10.1016/j.pec.2019.12.015] [Medline: 31901364]
63. Marasso D, Lupo C, Collura S, Rainoldi A, Brustio PR. Subjective versus objective measure of physical activity: A systematic review and meta-analysis of the convergent validity of the physical activity questionnaire for children (PAQ-C). Int J Environ Res Public Health 2021;18(7):3413 [FREE Full text] [doi: 10.3390/ijerph18073413] [Medline: 33806106]
64. Reilly JJ, Penpraze V, Hislop J, Davies G, Grant S, Paton JY. Objective measurement of physical activity and sedentary behaviour: review with new data. Arch Dis Child 2008;93(7):614-619. [doi: 10.1136/adc.2007.133272] [Medline: 18305072]
65. Montoye AH, Clevenger KA, Pfeiffer KA, Nelson MB, Bock JM, Imboden MT, et al. Development of cut-points for determining activity intensity from a wrist-worn ActiGraph accelerometer in free-living adults. J Sports Sci 2020;38(22):2569-2578. [doi: 10.1080/02640414.2020.1794244] [Medline: 32677510]
66. Hawlader MDH, Mozid N, Sharmin S, Monju IH, Ahmed SB, Sarker W, et al. The art of forming habits: applying habit theory in changing physical activity behaviour. J Public Health (Berl.) 2022;31(12):2045-2057. [doi: 10.1007/s10389-022-01766-4]
67. Riccio MT, Shrout PE, Balcetis E. Interpersonal pursuit of intrapersonal health goals: Social cognitive–motivational mechanisms by which social support promotes self‐regulatory success. Social & Personality Psych 2019;13(10):e12495. [doi: 10.1111/spc3.12495]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DBCT: dyadic behavior change technique
JITAI: just-in-time adaptive intervention
MVPA: moderate-to-vigorous physical activity
OSF: Open Science Framework

Edited by L Buis; submitted 07.Oct.2024; peer-reviewed by M vanDellen, J Novak, LO Wilhelm, H Namba; comments to author 20.Jun.2025; accepted 15.Nov.2025; published 27.Jan.2026.

Please cite as:
Höhener PS, Tobias R, Allen JM, Küng P, Scholz U
Effectiveness of Smartphone-Based Dyadic Interventions to Increase Physical Activity in Romantic Couples: Microrandomized Trial
JMIR Mhealth Uhealth 2026;14:e67136
URL: https://mhealth.jmir.org/2026/1/e67136
doi:10.2196/67136
PMID:

Individualized Treatment Effects of a Digital Smoking Cessation Intervention Among Individuals Looking Online for Help: Secondary Analysis of a Randomized Controlled Trial

Joel Crawford[1], PhD; Jenny Blomqvist[1], BSc; Katarina Ulfsdotter Gunnarsson[1], PhD; Preben Bendtsen[1,2], MD; Marcus Bendtsen[1], PhD

[1]Department of Health, Medicine and Caring Services, Linköping University, Linköping, Sweden

[2]Department of Medical Specialist, Motala Hospital, Motala, Sweden

Corresponding Author:
Joel Crawford, PhD
Department of Health, Medicine and Caring Services, Linköping University, Linköping, Sweden

Abstract

Background: Smoking is a leading cause of mortality and morbidity worldwide. Efforts to reduce smoking prevalence have used SMS text message–based interventions, which typically send participants a series of short, informational, motivational, and practical messages over a set period. Evidence highlights the efficacy of using this approach to support smoking cessation, with such trials typically reporting the average treatment effects, in which causal inference is made regarding the average effect of a treatment on a heterogeneous sample. Nonetheless, using this approach to assessing treatment effects means we are unable to account for individual factors that impact the effectiveness of a treatment on outcomes, such as age, gender, and genetics.

Objective: This study aimed to estimate the individualized effects of an SMS text message–based smoking cessation intervention to ascertain which individuals benefited the most and least during an effectiveness trial.

Methods: Data from a randomized controlled trial including 1012 adults from the Swedish general population were used. The trial assessed the effects of an SMS text message–based intervention, NEXit (Nicotine Exit), that aimed to change behavior by increasing the importance of change, boosting knowledge on how to change, and instilling confidence for change. Outcomes assessed in the trial were prolonged abstinence and point prevalence of smoking cessation. Individualized treatment effects were modeled using baseline factors (demographics, psychosocial variables, and past behavior) to study who benefited the most and least from the intervention.

Results: For prolonged abstinence, there was evidence of heterogeneous effects, with those benefiting the most from NEXit being older adults, female participants, individuals with high confidence in their ability to quit, and those who believed that quitting was important. For point prevalence abstinence, older individuals and those reporting high confidence in the ability to quit, the importance of quitting, and knowledge for change benefited the most. For both outcomes, individuals who reported smoking for a longer duration and smoking more at baseline benefited less.

Conclusions: The results demonstrate how individuals respond differently to an SMS text message–based smoking cessation intervention. This provides an insight into who benefits the most and least from the intervention in terms of demographics, baseline characteristics, and behaviors. The study highlights which individuals need to be specifically targeted and/or have content developed to suit their individual needs to further reduce the prevalence of smoking.

Trial Registration: ISRCTN Registry ISRCTN13455271; http://www.isrctn.com/ISRCTN13455271

(JMIR Mhealth Uhealth 2026;14:e63578) doi:10.2196/63578

KEYWORDS

individualized treatment effects; randomized controlled trial; text message–based smoking cessation intervention; smoking cessation; smoking; Sweden; behavior change; text messages; smoker

Introduction

Smoking remains a leading cause of morbidity and mortality across the globe, with recent figures highlighting approximately 8 million deaths worldwide and 200 million disability-adjusted life years attributed to smoking [1,2]. In Sweden, the prevalence of smoking within the general population has consistently decreased over the last 2 decades, with the Public Health Agency of Sweden reporting that prevalence rates dropped to 6% for women and 5% for men in 2022 [3]. Despite this, smoking remains a leading cause of mortality and morbidity within Sweden [4] and is more prevalent in groups with lower

socioeconomic status [5]. To stem the associated morbidity and mortality and reduce health inequalities, effective interventions aimed at reducing the prevalence of smoking within the general population are needed.

Current efforts to reduce smoking prevalence have used SMS text message–based interventions, which typically comprised a series of short messages sent to participants over a designated time period (usually 2-3 mo) [6,7]. The aim of these interventions is to provide motivation to quit, support for quitting, and reinforcement of the decision to quit [7]. A series of meta-analyses have highlighted the effectiveness of SMS text message–based smoking cessation interventions, and all concluded positive effects were observed for both self-reported and biochemically verified abstinence [6-10].

In Sweden, a recent randomized controlled trial (RCT) estimated the effectiveness of receiving a SMS text message–based smoking cessation intervention against existing public support (ie, referral to the national Quitline and website, slutarokalinjen.se) [11]. The trial targeted the general population using the NEXit (Nicotine Exit) protocol [12]. Effect estimates were made at 3 and 6 months post randomization for both prolonged abstinence (ie, the Russell standard definition of not having smoked more than 5 cigarettes in the past 8 wk) and point prevalence of abstinence. At 3 months post randomization, the odds ratio (OR) of prolonged abstinence was 2.15 (95% compatibility interval [CoI] 1.51-3.06), and at 6 months, the OR of prolonged abstinence was 2.38 (95% CoI 1.62-3.57). For point prevalence of abstinence, the OR at 3 months was 1.70 (95% CoI 1.18-2.44), and at 6 months, it was 1.49 (95% CoI 1.03-2.14).

While the findings suggest the intervention is effective for smoking cessation, the results of the RCT were assessed at the group level, meaning we are unable to account for how individual differences impact outcomes. This is pertinent, as not everyone in the intervention group successfully quit smoking, meaning that some individuals benefited more from the intervention than others. Individual differences in terms of fixed (eg, age, ethnicity, gender) and modifiable (eg, socioeconomic status, psychological) health factors have been shown to impact abstinence rates [13-16], which can potentially account for why some individuals benefit more from treatment than others.

Currently, Sweden has a national target of becoming a “smoke-free” state (ie, <5% prevalence in the general population) [17]. To achieve this target, smoking cessation support such as the NEXit protocol needs to become specifically targeted to those who continue to smoke tobacco despite policy measures and interventions. The aim of this study is to ascertain who benefited the most and least in the recent NEXit trial [11] through a secondary analysis of the group-level data, which estimated the average treatment effect of NEXit in the Swedish general population. We aim to achieve this by assessing how baseline factors influence outcomes of the trial at the individual level.

Methods

Study Design

The current study investigated the individualized effects of the NEXit protocol using data from a 2-arm, parallel-group (1:1) RCT. The trial was registered in the ISRCTN registry on December 3, 2020 (ISRCTN13455271), and a protocol and an analysis plan were published prospectively [12].

Participants and Settings

A total of 1012 participants from the Swedish general public were recruited from 2 settings: (1) online advertisements, in which individuals clicked on a link that took them to an information page with instructions for signing up, and (2) primary health care units in southern Sweden, which used printed media (leaflets and posters) containing instructions for signing up. Individuals were required to be aged 18 years or older, have access to a mobile phone, and smoke at least one cigarette per week to participate in the trial. Regardless of medium, those interested in participating sent an SMS text message to a dedicated number and received a reply within 5 minutes with a link to the study information and the informed consent form. After providing consent, participants were asked to complete a baseline survey, which was used to assess eligibility. Immediately afterward, eligible participants were randomly assigned to 1 of 2 study groups.

Ethical Considerations

The trial received ethical approval from the Swedish Ethical Review Authority on June 16, 2020 (Dnr 2020 - 01427). Participants provided informed consent prior to the commencement of the main trial, which included consent for their deidentified data to be analyzed and published. Trial data are deidentified, and participants in the main trial did not receive compensation.

Interventions

Both the control and intervention groups were offered treatment as usual and were thus not prohibited from using publicly available smoking cessation support aids. Treatment as usual was referral to a national smoking cessation helpline and website; those recruited from primary care units also had the option of meeting with a nurse or smoking cessation counselor. Additionally, participants in the intervention group were given access to the NEXit SMS text messaging intervention. The intervention had 2 fairly similar versions: a general version and a version for those undergoing elective surgery (within 3 mo of signing up). The intervention consists of a 1- to 4-week motivational phase in which participants were encouraged to set a date on which to quit, followed by a 12-week program consisting of 157 cessation support messages. Messages were typically informational and encouraging, and they provided tips and techniques for creating a smoke-free environment and lifestyle. These included distraction techniques, tips to avoid weight gain, tips to cope with cravings, tips on how to avoid triggers, tips to self-regulate, and practical tips to help restructure the physical environment. The design of the messages was informed by intervention mapping [18] and the behavior change wheel [19], with various behavior change techniques being

utilized, including 1.2 problem solving, 1.4 action planning, 5.1 information about health consequences, 5.2 salience of consequences, 8.2 behavioral substitution, 8.4 habit reversal, 9.1 credible source, 10.9 self-reward, 12 restricting the physical environment, 12.3 avoidance to cues, and 12.4 distraction. Messages were sent daily to participants, and the elective surgery version contained additional messages regarding the impact of smoking on surgery and recovery. Participants could also text and request additional support if they had relapsed or were experiencing cravings.

Outcomes and Measures

The primary outcomes of the trial were assessed at 3 and 6 months post randomization:

1. Prolonged abstinence, using the Russell standard definition of not having smoked more than 5 cigarettes in the past 8 weeks at the 3-month follow-up and in the past 5 months at the 6-month follow-up
2. Point prevalence, defined as not smoking any cigarettes during the past 4 weeks at both follow-up intervals

Baseline factors assessed in this secondary analysis of individualized effects include factors that are associated with cessation, including demographics (age and gender) and smoking behaviors (number of years smoking, amount of cigarettes smoked per week, use of snus [akin to snuff tobacco], number of past quit attempts, and nicotine dependence score using the Fagerström scale, which uses 6 items to assess intensity of physical addiction, including quantity and frequency of smoking) [13,20-24]. The recruitment setting (online or primary care unit) and whether participants were undergoing elective surgery were also assessed. Psychosocial variables (importance of quitting, confidence in the ability to quit, and knowledge on how to quit) were assessed. These factors are reflective of the behavioral and control beliefs that drive behavior, as outlined in expectancy-value models of health [25,26]. Extant literature highlights that the demographic factors (age and gender) are associated with other baseline factors used in our modeling [13,20-22].

Randomization and Blinding

Randomization was completed using a computer-generated random sequence, stratified according to which version of the intervention was appropriate (general or surgery). Block randomization, using block sizes of 2 and 4, ensured equal numbers of participants were allocated to the groups. Participants received notification of group allocation via SMS text message. Research personnel were blinded to allocation, and all procedures of the study were automated, except follow-up calls to nonresponders. For a full description of all study procedures, see Bendtsen et al [12].

Statistical Model

The observed outcome (Y) was modeled as a function of group allocation (G) and baseline variables (X) to estimate the individualized treatment effect $\delta(xi)=P(Yi|G=1,X=xi)-P(Yi=|G=0,X=xi)$, following the approach by Hoogland et al [27]. Here, δxi represents the difference between the predicted probability of smoking cessation (Yi) for an individual (i) with baseline factors xi, if they had access to the intervention (G=1) versus if they did not (G=0). Individualized treatment effects were modeled for each participant as a function of baseline factors to study which individuals had the most and least benefit from the intervention.

Smoking cessation outcomes at 3 and 6 months post randomization were modeled using multilevel logistic regression with a time-by-group interaction term. Available data were used to estimate the model and predict outcomes for all randomized participants. The models included participant-level adaptive intercepts, covariates for baseline factors, and interactions between baseline factors and group allocation. Bayesian inference was used to estimate the parameters of the models [28]. To promote parsimonious models, priors that shrunk estimates toward the null for baseline factor interaction coefficients were used, specifically Cauchy priors centered at the null with a standard normal hyperprior for the scale parameter [28]. We used Student t test priors centered at 0, with 3 df and a scale of 2.5 for noninteraction coefficients and standard normal priors for adaptive and fixed intercepts. Individualized treatment effects were calculated from posterior predictive draws, resulting in posterior distributions over treatment effects for each individual. Medians of these posterior distributions were used as point estimates of δxi.

Linear regression was used to study associations between standardized estimated individualized effects and baseline factors. Student t test priors centered at 0 with 3 df and a scale of 2.5 for all coefficients in these models. Medians of posterior distributions of covariates were used as point estimates and created 95% CoI using the 2.5 and 97.5 percentiles of the posterior distributions. Estimation was done using CmdStan version 2.33.0 (Stan Project), which is a shell interface for Stan [29]. Plots were generated using the ggplot2 library in R.

Results

Participant Retention and Missing Data

Table 1 presents baseline factors for the 1012 participants recruited to the trial. At the 3-month follow-up interval, data were available for 66% of participants in the intervention group and 69% of participants in the control group. The corresponding proportions for the 6-month interval were 64% for both groups. These data were used to estimate the multilevel logistic regression models, which were subsequently used to estimate individualized treatment effects for all randomized participants. As reported in the primary findings of the trial [11], there was evidence that older participants were more likely to respond to follow-up at the 6-month interval, but no other associations were found between response and baseline factors. The CONSORT (Consolidated Standards of Reporting Trials) flowchart and checklist are presented in Figure 1 and Checklist 1.

Table . Baseline factors of participants.

	Total (N=1012)	Intervention (n=505)	Control (n=507)
Woman[a], n (%)	820 (81)	406 (80.4)	414 (81.7)
Age (years), mean (SD)	45.4 (14.0)	45 (13.9)	45.7 (14.1)
Years of smoking, mean (SD)	25.3 (14.6)	24.7 (14.3)	25.9 (14.9)
Daily smokers (vs weekly smokers), n (%)	981 (96.9)	489 (96.8)	492 (97)
Cigarettes smoked per week, mean (SD)	101 (46.2)	101.4 (47.3)	100.6 (45.1)
Use of snus[b], n (%)			
Daily	63 (6.2)	27 (5.3)	36 (7.1)
Weekly or monthly	90 (8.9)	45 (8.9)	45 (8.9)
No	859 (84.9)	433 (85.7)	426 (84)
Fagerström test for nicotine dependence, mean (SD)	5 (2.2)	5 (2.2)	5 (2.2)
Quit attempts[c], mean (SD)	7.2 (13.7)	7.0 (12.7)	7.5 (14.6)
Cessation counseling experience, n (%)			
Yes, now	32 (3.2)	13 (2.6)	19 (3.7)
Yes	192 (19)	95 (18.8)	97 (19.1)
No	788 (77.9)	397 (78.6)	391 (77.1)
Used quit smoking helpline, n (%)	13.6 (138)	14.5 (73)	12.8 (65)
Importance of quitting[d], mean (SD)	9.4 (1.3)	9.4 (1.3)	9.5 (1.2)
Confidence in ability to quit[d], mean (SD)	6.2 (2.5)	6.3 (2.5)	6.2 (2.6)
Knowledge of how to quit[d], mean (SD)	5.5 (2.6)	5.5 (2.7)	5.5 (2.5)
Elective surgery[e], n (%)	63 (6.2)	31 (6.1)	32 (6.3)

[a]The baseline questionnaire included a category “Other”; however, it was not chosen by any participant.

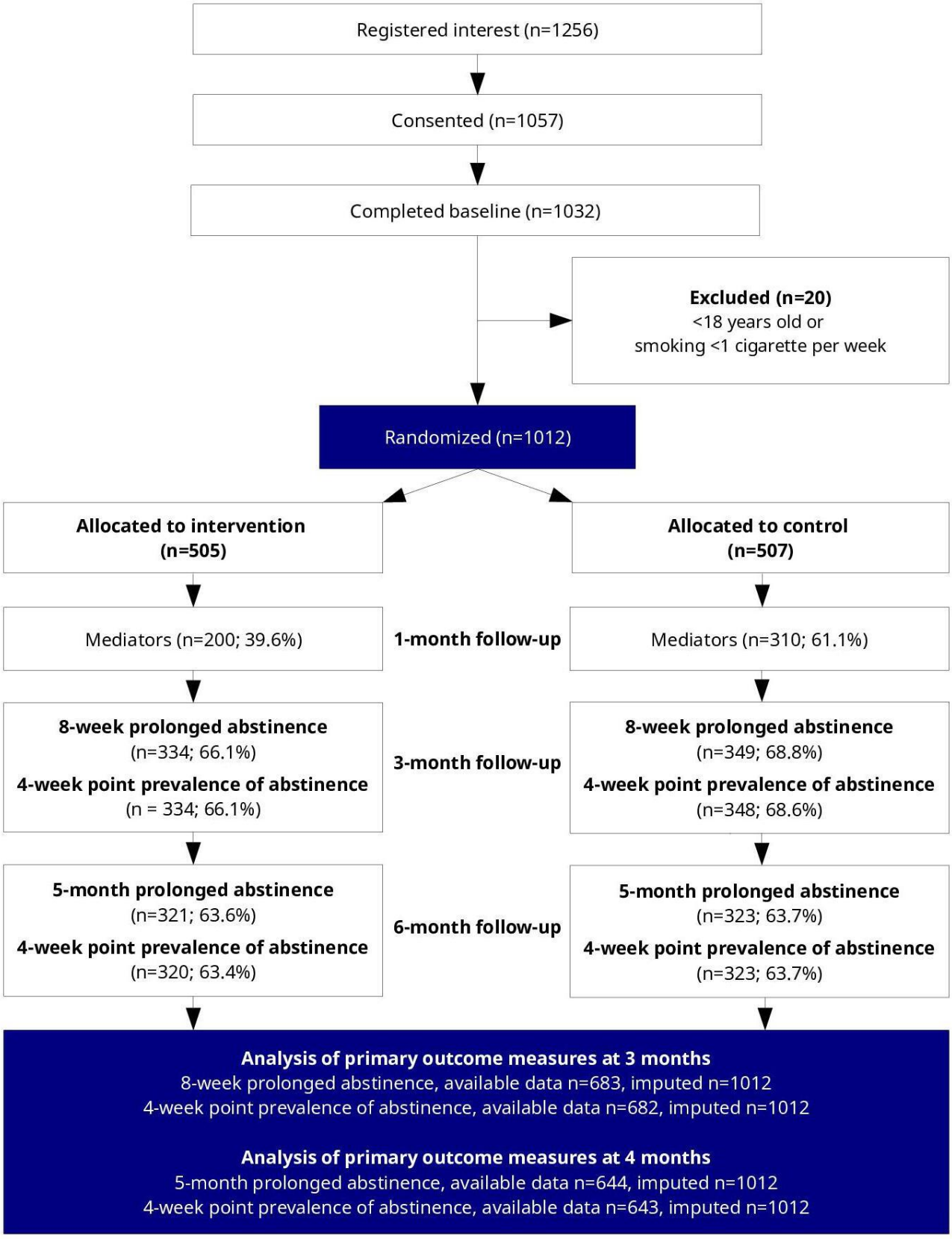
[b]Snus is a moist oral tobacco product, which is common in Sweden, sometimes translated as snuff.

[c]Participants were asked about the lifetime number of quit attempts.

[d]Three single-item measures were used to assess the importance, confidence, and know-how regarding smoking cessation. Responses ranged from 0 to 10, with 10 representing the highest agreement (ie, very important, very confident, very knowledgeable). The same items were used at follow-up as hypothesized mediators of effects.

[e]Participants self-reporting that they had planned elective surgery received additional intervention materials regarding smoking and surgery.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart.



Baseline Factors and Individualized Treatment Effects

In Figure 2, the distribution of effects for 8-week or 5-month abstinence is presented for both the 3- and 6-month follow-up intervals, and Figure 3 provides the distribution of effects for point prevalence abstinence for both follow-ups. The figures show the effects of NEXit at the individual level, expressed as differences in probability (percentage points) of smoking cessation between having received the intervention and the control. From the figures, it is evident that most participants were predicted to benefit from the intervention; however, more

so in terms of prolonged abstinence at the 6-month follow-up interval. The mean individualized effect for point prevalence at the 3-month interval was 3.9 (SD 5.3) percentage points and 2.7 (SD 3.9) percentage points at the 6-month interval. For prolonged abstinence, the mean individualized effect was 7.5 (SD 8.5) percentage points at 3 months and 7.8 (SD 10.7) percentage points at 6 months. This indicates that the intervention was most effective in producing prolonged abstinence, but more importantly, that individualized effects were overdispersed and showed heterogeneity of effects in the population.

Figure 2. Distribution of effects for prolonged abstinence.

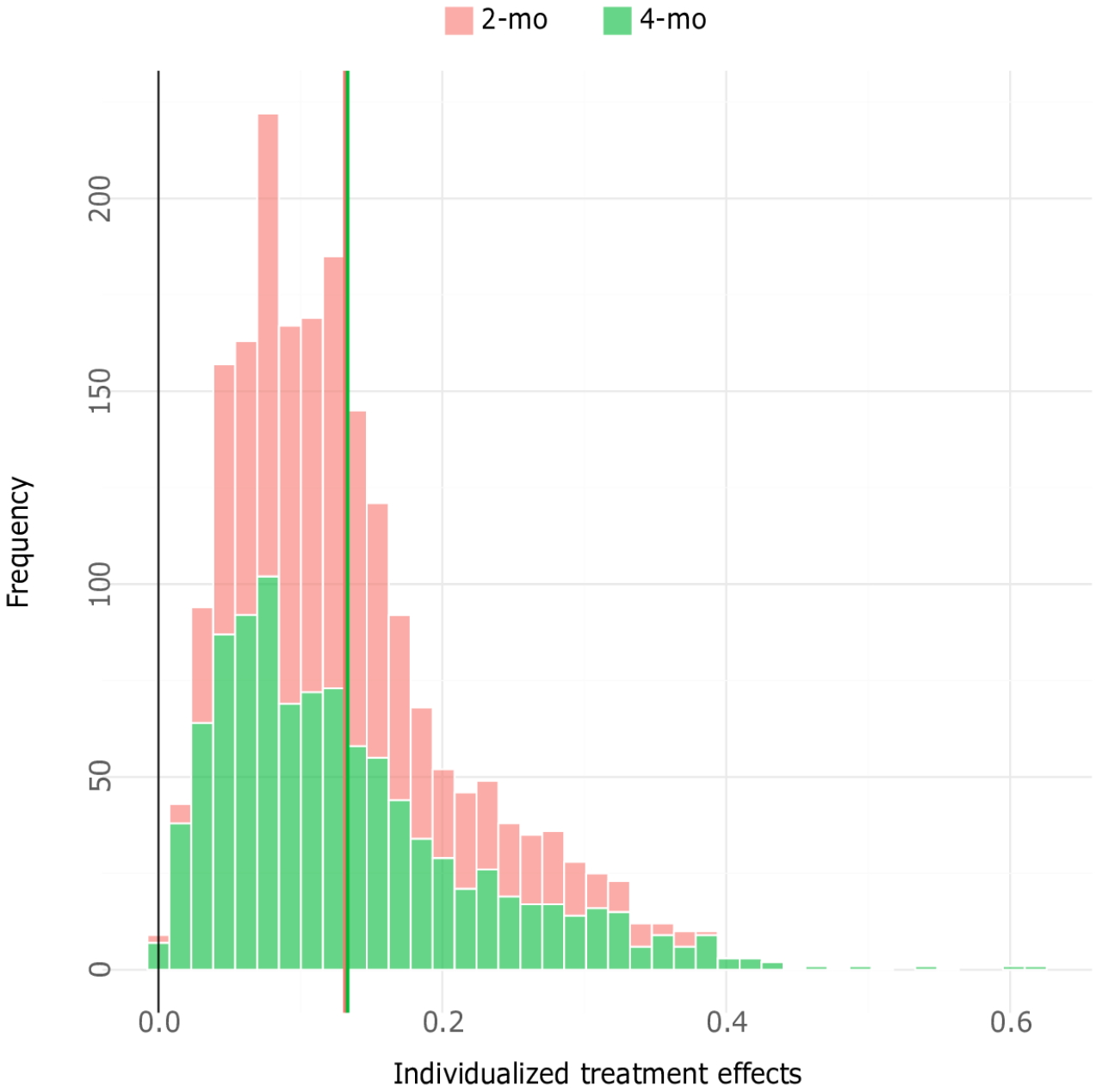
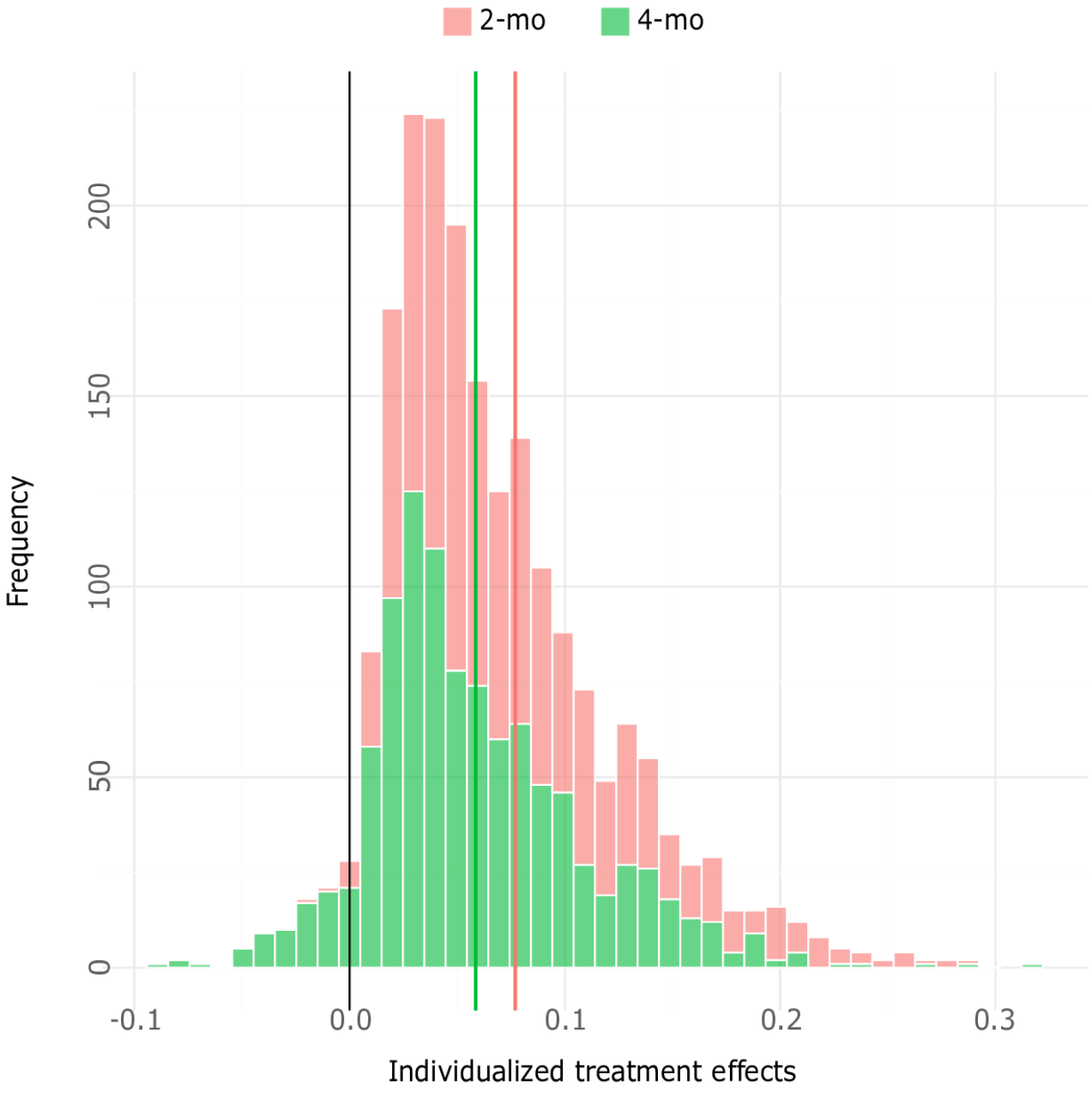


Figure 3. Distribution of effects for point prevalence abstinence.



Prolonged Abstinence

In Table 2, associations between baseline factors and standardized individualized treatment effects on prolonged abstinence are presented for both follow-up intervals. For both time points (3 and 6 mo), there was marked evidence for positive associations between individualized effects and gender, age, importance, and confidence. Meanwhile, there was marked evidence for negative associations between individualized effect and number of cigarettes smoked at baseline and years of smoking. Associations between individualized effects and recruitment setting, and past professional cessation support, were found for the 6-month follow-up period.

Table . Baseline factors and standardized individualized effects on prolonged abstinence.

	8 wk		5 mo	
	Est[a], OR[b] (95% CoI[c])	Prob[d] (%)	Est[a], OR (95% CoI)	Prob[d] (%)
Man vs woman	0.34 (0.19 to 0.49)	>99.9	0.18 (0.03 to 0.32)	99.1
Age	0.02 (0.01 to 0.03)	>99.9	0.02 (0.02 to 0.03)	>99.9
Importance	0.09 (0.05 to 0.14)	>99.9	0.1 (0.05 to 0.14)	>99.9
Confidence	0.05 (0.03 to 0.08)	>99.9	0.06 (0.03 to 0.08)	>99.9
Know-how	0.01 (-0.01 to 0.04)	83.7	0.02 (0.0 to 0.05)	97.6
Cigarettes smoked	–0.01 (–0.01 to –0.01)	>99.9	–0.01 (–0.01 to –0.004)	>99.9
Years smoking	–0.01 (–0.02 to –0.004)	99.8	–0.02 (–0.03 to –0.01)	>99.9
Quit attempts	–0.01 (–0.01 to –0.004)	>99.9	0.01 (0.006 to 0.014)	>99.9
Snus, weekly or monthly vs daily	–0.08 (–0.37 to 0.21)	70.5	–0.11 (–0.40 to 0.18)	76.0
Snus, none vs daily	–0.11 (–0.35 to 0.12)	83.1	0.02 (–0.21 to 0.25)	56.1
Fagerström score	0.01 (–0.02 to 0.04)	72.9	0.02 (–0.01 to 0.06)	89.4
Elective surgery	0.63 (0.40 to 0.86)	>99.9	0.50 (0.27 to 0.73)	>99.9
Online vs primary care	–0.09 (–0.37 to 0.20)	71.7	–0.46 (–0.74 to –0.17)	99.9
Quitline	–0.03 (–0.20 to 0.13)	64.2	0.14 (–0.03 to 0.30)	95.1
Past cessation support	–0.12 (–0.27 to 0.03)	93.5	–0.21 (–0.36 to –0.06)	99.7
Current cessation support	0.24 (–0.08 to 0.56)	92.6	0.17 (–0.15 to 0.48)	84.6

[a]Est: median of the posterior distribution of standardized individual effects with 95% CoIs.

[b]OR: odds ratio.

[c]CoI: compatibility interval.

[d]Prob: proportion of the posterior distribution above or below the null in the direction of the median.

Point Prevalence Abstinence

In Table 3, associations between baseline factors and standardized individualized treatment effects on point prevalence are presented for both follow-up intervals. For both time points, there was marked evidence of positive associations between individualized effect and age, importance, confidence, know-how, and elective surgery at the 6-month interval. Negative associations between individualized effects and the number of cigarettes smoked at baseline and years of smoking were found.

Table . Baseline factors and standardized individualized effects on point prevalence.

	3 mo		6 mo	
	Est[a], OR[b] (95% CoI[c])	Prob[d] (%)	Est[a], OR (95% CoI)	Prob[d] (%)
Man vs Woman	0.05 (–0.10 to 0.20)	73.7	0.011 (–0.13 to 0.16)	55.7
Age	0.03 (0.02 to 0.03)	>99.9	0.02 (0.011 to 0.03)	>99.9
Importance	0.05 (0.01 to 0.10)	98.5	0.07 (0.03 to 0.12)	99.9
Confidence	0.05 (0.02 to 0.07)	>99.9	0.04 (0.02 to 0.07)	>99.9
Know-How	0.02 (–0.004 to 0.04)	94.4	0.02 (–0.001 to 0.05)	97.0
Cigarettes smoked	–0.01 (–0.01 to –0.01)	>99.9	–0.01 (–0.01 to –0.01)	>99.9
Years smoking	–0.02 (–0.03 to –0.02)	>99.9	–0.03 (–0.04 to –0.02)	>99.9
Quit attempts	–0.002 (–0.01 to 0.002)	84.6	0.002 (–0.002 to 0.01)	85.7
Snus, weekly or monthly vs daily	–0.03 (–0.32 to 0.26)	57.2	–0.43 (–0.71 to –0.14)	99.9
Snus, none vs daily	–0.12 (–0.35 to 0.11)	84.4	–0.13 (–0.91 to 0.65)	63.7
Fagerström score	–0.02 (–0.05 to 0.01)	86.6	0.01 (–0.02 to 0.04)	75.1
Elective surgery	0.20 (–0.03 to 0.43)	95.7	0.37 (0.15 to 0.60)	99.9
Online vs primary care	0.19 (–0.09 to 0.47)	90.4	0.01 (–0.27 to 0.29)	51.5
Quitline	0.09 (–0.08 to 0.25)	84	0.01 (–0.16 to 0.18)	53.9
Past cessation support	–0.04 (–0.19 to 0.11)	71.7	–0.04 (–0.19 to 0.11)	70.7
Current cessation support	–0.12 (–0.45 to 0.21)	76.1	–0.13 (–0.45 to 0.19)	78.6

[a]Est: Median of the posterior distribution of standardized individual effects with 95% compatibility intervals (CoI).

[b]OR: odds ratio.

[c]CoI: compatibility interval.

[d]Prob: Proportion of the posterior distribution above or below the null in the direction of the median.

Interactions

To assess whether the heterogeneous main associations for age and gender were qualified by interactions, models with added interaction terms were estimated. The only interaction for which there was marked evidence was between gender and importance to quit smoking, which was associated with individualized effects of the intervention on prolonged abstinence. This suggests that the main association of gender for this outcome was accounted for by its interaction with importance. Please see Table 4 for estimates of the interaction and Figure 4 for plots illustrating these interactions.

Table . Interaction associations of gender and baseline importance to quit smoking on prolonged abstinence.

	8 wk		5 mo	
	Est[a], OR[b] (95% CoI[c])	Prob[d] (%)	Est[a], OR (95% CoI)	Prob[d] (%)
Man vs woman	0.004 (–0.12 to 0.28)	56.6	0.0 (–0.22 to 0.18)	50
Importance	0.06 (0.011 to 0.11)	99.4	0.08 (0.03 to 0.13)	>99.9
Man vs woman × importance	0.03 (0.01 to 0.05)	98.4	0.02 (–0.004 to 0.05)	95.4

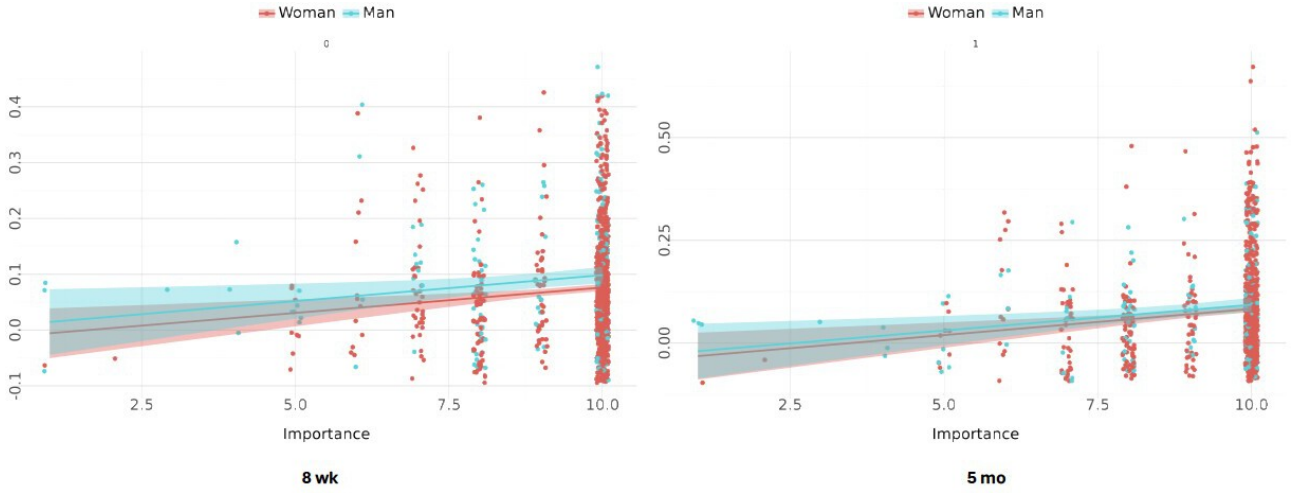
[a]Est: Median of the posterior distribution of standardized individual effects with 95% compatibility intervals (CoI).

[b]OR: odds ratio.

[c]CoI: compatibility interval.

[d]Prob: Proportion of the posterior distribution above or below the null in the direction of the median.

Figure 4. Visualization of interaction associations of gender and baseline importance to quit smoking on prolonged abstinence.



Discussion

Age

The current study assessed the individualized treatment effects of the recent NEXit SMS text message–based intervention [11]. We found an association between age and individualized effects (ie, as age increases beyond the mean value), with older participants benefiting more than younger participants for both prolonged and point prevalence abstinence. This finding contrasts with evidence comparing effects across trials, which highlights that effective interventions for the adult population in general have similar effects for younger adults [30]. This disparity may be a result of the content of the intervention, which specifically focused on support to restructure the physical environment rather than support for the social environment. This is important, as younger adults have a high prevalence of social smoking [20], which may account for this effect. Future interventions for this subsample could include aspects targeting social factors. Nonetheless, evidence also highlights that younger adults are more likely to initiate cessation compared to older adults and smoke less heavily [13,20]; therefore, the results of the current analysis are encouraging, as they suggest that the subsample who are less likely to quit and who smoke more benefit the most from NEXit. The results also suggest that older adults in Sweden are willing to engage with digital health interventions. This finding contrasts with evidence that suggests a "digital divide" exists between older and younger adults in terms of acceptance and uptake of digital health interventions [31]. The current findings suggest that SMS text message–based interventions may be particularly apt for older adults.

Confidence and Know-How

Baseline confidence was associated with individualized effects for both prolonged and point prevalence abstinence, suggesting that those who had greater confidence in their ability to quit benefited more from NEXit. This finding supports extant literature that highlights higher rates of self-efficacy (ie, confidence) are associated with a greater likelihood of successful smoking cessation [32-34]. The results for know-how varied, with no association found for the first time point for both prolonged and point prevalence abstinence; however, an association was found at the second time point for both outcome measures, with those displaying a higher rating of know-how benefiting more. It may be the case that participants had been lacking knowledge on how to quit and maintain abstinence over a short period of time, while they may have had knowledge on how to sustain abstinence over a longer period. This suggests that NEXit may be beneficial to those who are unsure of how to initially stop smoking and maintain the initial phase of abstinence.

Gender

Associations between both importance and gender with individualized effects were found, with those who deemed it of greater importance to quit benefiting more, while effects were more pronounced for men compared to women. The finding for importance supports extant literature that has assessed motivation to quit (ie, the reasons smokers consider it important to quit), which suggests that higher motivation increases the likelihood of cessation and predicts abstinence [34-37]. Results compared across effectiveness trials highlight that women are less likely to remain abstinent than men [22], and the findings of the current study support this suggestion. Nonetheless, the association between effect and gender dissipated when adding the interaction term of gender and importance to the model, suggesting that, between men and women who deemed it of greater importance to quit, the effects of NEXit were greater for men. Future studies should consider exploring why importance may not be a driving force for change among women but is so for men. It should be noted that the sample included more women than men, reflecting greater treatment seeking among women. The parent trial was pragmatic in the sense that it was an effectiveness trial recruiting participants in a way that mimicked how a real-world implementation could work (ie, online advertisements). Therefore, our findings should most conservatively be generalized to those seeking help online, and not the general population, where the prevalence of smoking is more balanced.

On the other hand, for point prevalence, there were no marked associations between effects or interactions with respect to gender. This suggests that rates of relapse (or lack of) may have been similar between women and men, which is in contrast to evidence suggesting that women are more likely to relapse than

men [21]. The content for craving provided by NEXit may have been equally effective for the needs of both genders; evidence suggests that affective factors such as mood, anger, and perceived stress predict relapse in women, while the motivation to reduce craving is a unique relapse predictor in men [38]. The support for craving had specific content relating to stress management, reframing perspectives, and motivation.

The results highlight that confidence and importance (especially for men) are potential focal considerations for smoking cessation. Confidence and importance are integral parts of the decision to quit smoking, especially for individuals who are contemplating or preparing for a change and who may feel ambivalent about quitting [39]. It may be that those who reported the importance of quitting and their confidence as high were better psychologically prepared to change behavior. This suggests that future interventions should target improvements in both factors.

Smoking and Number of Quit Attempts

Associations between the number of cigarettes smoked and years of smoking with individualized effects were found for both prolonged and point prevalence abstinence, with those who smoked less and for a shorter duration benefiting the most. This suggests that NEXit is less beneficial for those who smoke more and have done so for a longer duration. Theories of addiction posit that when an addiction progresses, such as nicotine dependence, control over consumption shifts from goal-directed behavior to automatic or habitual motivation [40], hence those who smoke more and for longer *may* have been driven more by habit than goal-directed behavior. While the content of NEXit did include advice on how to avoid smoking cues, this may not have been enough for these heavier, longer-term smokers. Providing greater support for identifying what triggers habitual smoking and advice on how to plan for encountering triggers or cues may be beyond the scope of an SMS text message–based intervention. Potentially, this could be better achieved by smoking cessation counseling or through the usage of an interactive digital smoking cessation intervention, whereby end users are encouraged to record how they are triggered, along with creating their own plans to overcome their triggers, which can then be sent as reminders.

For prolonged abstinence, an association between quit attempts and individualized effect was found at 8 weeks, with those having fewer attempts at baseline benefiting more. The underlying action of NEXit may account for this finding, with the intervention theorized to be driven by instilling confidence in ability and raising the importance of quitting. Evidence highlights that self-efficacy (ie, confidence) and motivation (ie, importance) are key factors in smoking cessation for individuals who are less likely to make a quit attempt [37]. It may be that the mediating effects of confidence and importance on prolonged abstinence had a greater impact on this group than on those with greater recorded quit attempts at baseline. Nonetheless, at the 5-month follow-up, an individualized effect on prolonged abstinence was found for those reporting more quit attempts at baseline. This effect is potentially driven by these individuals building more confidence and gaining knowledge for quitting over the course of the trial.

Degree of Dependence

Finally, for the dependence score, only an association with individualized effect for point prevalence at 3 months was found, with those reporting less dependence benefiting more. No other associations with dependence were observed. Dependence as a factor in cessation and abstinence is well established; the likelihood of quitting is much greater at low levels of dependence, in comparison to high levels [41]. The disparity in our findings with the literature may be due to the methods used to assess dependency. We used a self-report measure, whereas many studies use biochemical measures of dependence [42]. While the Fagerström scale has been shown to predict abstinence across trials [24], self-report measures are more prone to biases in comparison to objective measures such as a saliva cotinine test. While future studies should consider both self-reports and biochemical measures, it may not be feasible for the digital modality. Furthermore, evidence highlights selection biases present in saliva testing [43], while recommendations suggest self-report measures are more appropriate for general population studies [43]. Nonetheless, our results for dependence could be interpreted to mean that, by and large, NEXit is equally effective regardless of dependence severity.

Limitations

The results provide insight into the individualized effects of the recent NEXit trial; however, they should be interpreted with some considerations. Typically in RCTs, the randomized component is tested, rather than a comparison across treatments, meaning individuals may adhere to or engage in an intervention at their own preference. Therefore, the results should be considered in terms of having access to the intervention rather than using it. Indeed, the aim of the trial was to estimate the effectiveness of providing a digital smoking cessation intervention; therefore, all intercurrent events were acceptable. We do note that since we did not collect data on the use of other interventions throughout the trial period, we cannot investigate if the NEXit intervention increased the utilization of other treatments.

Another limitation of this study is that a statistical model was used to predict and contrast outcomes. This means that the estimated effects are not based entirely on observed data but on predicted data. These predictions may have considerable uncertainty that is not readily quantifiable, as no individuals received both conditions; thus, no individual contrasts exist against which predictions can be compared. The result of this is that when comparing individualized effects against baseline factors, there is uncertainty that is not accounted for. More research is needed to develop methods to estimate the uncertainty of the statistical approach taken [27].

It should be noted that the relationships between baseline characteristics and individualized effects are associational and should not be interpreted as causal. Confounding and collider bias can substantially alter estimated associations. Further, the current study investigated several associations between baseline characteristics and individualized effects. The Bayesian inference paradigm informs which associations are stronger and weaker relative to one another and which associations have the

relatively strongest evidence, conditional on the data and prior knowledge. By using Student *t* test priors centered at 0, there was weak shrinkage induced in these estimations to protect against overestimation. Thus, the result of this study is all the associations presented and not only those that pass a prespecified threshold of significance (eg, null-hypothesis testing). In all, the findings presented in this study should be seen as exploratory, and future studies should look to confirm the associations as causal.

Conclusions

In conclusion, the results of this study highlight that individuals participating in the RCT of the NEXit intervention had heterogeneous outcomes, which were not discernible from the primary analysis of the RCT that compared the intervention and control groups [11]. The intervention was most effective among men, older individuals, those who smoked less at baseline and for a shorter duration, and those who had confidence in their ability and deemed it important to quit. The results have the potential to help refine smoking cessation interventions so that we can target those who persist in smoking despite current intervention efforts and policy measures.

Acknowledgments

Artificial intelligence chatbots were not used in any part of the manuscript writing.

Funding

This study was conducted under the auspices of a grant from The Swedish Cancer Society (Cancerfonden, 23 2888; principal investigator: MB). The parent trial was funded by Region Östergötland (Forsknings-ALF, Dnr: LIO-896081). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the General Data Protection Regulation but are available upon reasonable request to marcus.bendtsen@liu.se.

Authors' Contributions

Data analysis: MB
Data collection: JB, MB
Editing manuscript: KUG, PB, MB
Writing – lead and editing: JC

Conflicts of Interest

MB and PB own a private company (Alexit AB) that maintains and distributes evidence-based lifestyle interventions to be used by the public and in health care settings. Alexit AB played no role in developing the intervention, study design, data analysis, data interpretation, or writing of this report. Services developed and maintained by Alexit AB were used for sending SMS text messages and data collection. All other authors declare no conflicts of interest.

Checklist 1
CONSORT checklist.
[PDF File, 239 KB - mhealth_v14i1e63578_app1.pdf]

References

1. Reitsma MB, Kendrick PJ, Ababneh E, GBD 2019 Tobacco Collaborators. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990-2019: a systematic analysis from the Global Burden of Disease Study 2019. Lancet 2021 Jun 19;397(10292):2337-2360. [doi: 10.1016/S0140-6736(21)01169-7] [Medline: 34051883]
2. Tobacco. World Health Organization. URL: https://www.who.int/health-topics/tobacco#tab=tab_1 [accessed 2026-01-06]
3. Vuxnas bruk av tobaks- och nikotinprodukter [Article in Swedish]. Folkhälsomyndigheten. URL: https://www.folkhalsomyndigheten.se/livsvillkor-levnadsvanor/andts/andts-anvandning-och-ohalsa/anvandning-och-omfattning-av-andts-i-befolkningen/anvandning-av-tobaks-och-nikotinprodukter/vuxnas-bruk-av-tobaks--och-nikotinprodukter/ [accessed 2025-01-06]
4. Tobacco and nicotine products. Public Health Agency of Sweden. URL: https://www.folkhalsomyndigheten.se/the-public-health-agency-of-sweden/living-conditions-and-lifestyle/andtg/tobacco/ [accessed 2026-01-06]

5. Molarius A, Hellstrand M, Engström S. Social differences in who receives questions and advice about smoking habits when visiting primary care - results from a population based study in Sweden in 2012. Prev Med Rep 2017 Mar;5:236-240. [doi: 10.1016/j.pmedr.2016.12.016] [Medline: 28127526]
6. Spohr SA, Nandy R, Gandhiraj D, Vemulapalli A, Anne S, Walters ST. Efficacy of SMS text message interventions for smoking cessation: a meta-analysis. J Subst Abuse Treat 2015 Sep;56:1-10. [doi: 10.1016/j.jsat.2015.01.011] [Medline: 25720333]
7. Scott-Sheldon LAJ, Lantini R, Jennings EG, et al. Text messaging-based interventions for smoking cessation: a systematic review and meta-analysis. JMIR Mhealth Uhealth 2016 May 20;4(2):e49. [doi: 10.2196/mhealth.5436] [Medline: 27207211]
8. Mason M, Ola B, Zaharakis N, Zhang J. Text messaging interventions for adolescent and young adult substance use: a meta-analysis. Prev Sci 2015 Feb;16(2):181-188. [doi: 10.1007/s11121-014-0498-7] [Medline: 24930386]
9. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y, Dobson R. Mobile phone text messaging and app-based interventions for smoking cessation. Cochrane Database Syst Rev 2019 Oct 22;10(10):CD006611. [doi: 10.1002/14651858.CD006611.pub5] [Medline: 31638271]
10. Zhou X, Wei X, Cheng A, et al. Mobile phone-based interventions for smoking cessation among young people: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023 Sep 12;11:e48253. [doi: 10.2196/48253] [Medline: 37706482]
11. Blomqvist J, Gunnarsson KU, Bendtsen P, Bendtsen M. Effects of a text messaging smoking cessation intervention amongst online help-seekers and primary health care visitors: findings from a randomised controlled trial. BMC Med 2023 Oct 4;21(1):382. [doi: 10.1186/s12916-023-03073-5] [Medline: 37794399]
12. Bendtsen M, Thomas K, Linderoth C, Bendtsen P. Effects of a text messaging smoking cessation intervention among online help seekers and primary health care visitors in Sweden: protocol for a randomized controlled trial using a Bayesian group sequential design. JMIR Res Protoc 2020 Dec 3;9(12):e23677. [doi: 10.2196/23677] [Medline: 33269703]
13. Arancini L, Borland R, Le Grande M, et al. Age as a predictor of quit attempts and quit success in smoking cessation: findings from the International Tobacco Control Four-Country survey (2002-14). Addiction 2021 Sep;116(9):2509-2520. [doi: 10.1111/add.15454] [Medline: 33651412]
14. Nollen NL, Mayo MS, Sanderson Cox L, et al. Factors that explain differences in abstinence between Black and White smokers: a prospective intervention study. J Natl Cancer Inst 2019 Oct 1;111(10):1078-1087. [doi: 10.1093/jnci/djz001] [Medline: 30657926]
15. Kock L, Brown J, Hiscock R, Tattan-Birch H, Smith C, Shahab L. Individual-level behavioural smoking cessation interventions tailored for disadvantaged socioeconomic position: a systematic review and meta-regression. Lancet Public Health 2019 Dec;4(12):e628-e644. [doi: 10.1016/S2468-2667(19)30220-8] [Medline: 31812239]
16. Boudrez H. Psychological factors and long-term abstinence after smoking cessation treatment. J Smok Cessat 2009 May 1;4(1):10-17. [doi: 10.1375/jsc.4.1.10]
17. Hallengren L. Rökfritt sverige 2025 [Report in Swedish]. : Sveriges Riksdag; 2017 URL: https://data.riksdagen.se/fil/4E855498-EF93-426C-9420-E13DA6E88E18 [accessed 2026-01-06]
18. Eldredge LKB, Markham CM, Ruiter RA, Fernandez ME. Planning Health Promotion Programs: An Intervention Mapping Approach, 4th edition: Jossey-Bass Inc; 2016. URL: https://cris.maastrichtuniversity.nl/en/publications/planning-health-promotion-programs-an-intervention-mapping-approa/ [accessed 2026-01-06]
19. Miche S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions, 1st edition: Silverback Publishing; 2014. URL: https://www.behaviourchangewheel.com/about-book [accessed 2026-01-06]
20. Song AV, Ling PM. Social smoking among young adults: investigation of intentions and attempts to quit. Am J Public Health 2011 Jul;101(7):1291-1296. [doi: 10.2105/AJPH.2010.300012] [Medline: 21566040]
21. Swan GE, Ward MM, Carmelli D, Jack LM. Differential rates of relapse in subgroups of male and female smokers. J Clin Epidemiol 1993 Sep;46(9):1041-1053. [doi: 10.1016/0895-4356(93)90172-w] [Medline: 8263577]
22. Smith PH, Bessette AJ, Weinberger AH, Sheffer CE, McKee SA. Sex/gender differences in smoking cessation: a review. Prev Med 2016 Nov;92:135-140. [doi: 10.1016/j.ypmed.2016.07.013] [Medline: 27471021]
23. Gram IT, Antypas K, Wangberg SC, Løchen ML, Larbi D. Factors associated with predictors of smoking cessation from a Norwegian internet-based smoking cessation intervention study. Tob Prev Cessat 2022;8:38. [doi: 10.18332/tpc/155287] [Medline: 36382026]
24. Fagerström K, Russ C, Yu CR, Yunis C, Foulds J. The Fagerström test for nicotine dependence as a predictor of smoking abstinence: a pooled analysis of varenicline clinical trial data. Nicotine Tob Res 2012 Dec;14(12):1467-1473. [doi: 10.1093/ntr/nts018] [Medline: 22467778]
25. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
26. Maddux JE, Rogers RW. Protection motivation and self-efficacy: a revised theory of fear appeals and attitude change. J Exp Soc Psychol 1983 Sep;19(5):469-479. [doi: 10.1016/0022-1031(83)90023-9]
27. Hoogland J, IntHout J, Belias M, et al. A tutorial on individualized treatment effect prediction from randomized trials with a binary endpoint. Stat Med 2021 Nov 20;40(26):5961-5981. [doi: 10.1002/sim.9154] [Medline: 34402094]
28. Bendtsen M. A gentle introduction to the comparison between null hypothesis testing and Bayesian analysis: reanalysis of two randomized controlled trials. J Med Internet Res 2018 Oct 24;20(10):e10873. [doi: 10.2196/10873] [Medline: 30148453]

29. Stan Development Team. Stan documentation. Stan. 2023. URL: https://mc-stan.org/docs/ [accessed 2026-01-06]
30. Suls JM, Luger TM, Curry SJ, Mermelstein RJ, Sporer AK, An LC. Efficacy of smoking-cessation interventions for young adults: a meta-analysis. Am J Prev Med 2012 Jun;42(6):655-662. [doi: 10.1016/j.amepre.2012.02.013] [Medline: 22608385]
31. McDonough CC. The effect of ageism on the digital divide among older adults. Gerontol Geriatr Med 2016 Jun 16;2(1):1-7. [doi: 10.24966/GGM-8662/100008]
32. Gallus S, Cresci C, Rigamonti V, et al. Self-efficacy in predicting smoking cessation: a prospective study in Italy. Tob Prev Cessat 2023;9(April):15. [doi: 10.18332/tpc/162942] [Medline: 37125003]
33. Klemperer EM, Mermelstein R, Baker TB, et al. Predictors of smoking cessation attempts and success following motivation-phase interventions among people initially unwilling to quit smoking. Nicotine Tob Res 2020 Aug 24;22(9):1446-1452. [doi: 10.1093/ntr/ntaa051] [Medline: 32236417]
34. Boardman T, Catley D, Mayo MS, Ahluwalia JS. Self-efficacy and motivation to quit during participation in a smoking cessation program. Int J Behav Med 2005;12(4):266-272. [doi: 10.1207/s15327558ijbm1204_7] [Medline: 16262545]
35. Wee LH, West R, Bulgiba A, Shahab L. Predictors of 3-month abstinence in smokers attending stop-smoking clinics in Malaysia. Nicotine Tob Res 2011 Feb;13(2):151-156. [doi: 10.1093/ntr/ntq221] [Medline: 21186253]
36. Williams GC, Gagné M, Ryan RM, Deci EL. Facilitating autonomous motivation for smoking cessation. Health Psychol 2002 Jan;21(1):40-50. [doi: 10.1037/0278-6133.21.1.40] [Medline: 11846344]
37. Jardin BF, Carpenter MJ. Predictors of quit attempts and abstinence among smokers not currently interested in quitting. Nicotine Tob Res 2012 Oct;14(10):1197-1204. [doi: 10.1093/ntr/nts015] [Medline: 22387995]
38. Nakajima M, al'Absi M. Predictors of risk for smoking relapse in men and women: a prospective examination. Psychol Addict Behav 2012 Sep;26(3):633-637. [doi: 10.1037/a0027280] [Medline: 22352701]
39. DiClemente CC, Prochaska JO. Self-change and therapy change of smoking behavior: a comparison of processes of change in cessation and maintenance. Addict Behav 1982;7(2):133-142. [doi: 10.1016/0306-4603(82)90038-7] [Medline: 7102444]
40. Doñamayor N, Ebrahimi C, Arndt VA, Weiss F, Schlagenhauf F, Endrass T. Goal-directed and habitual control in human substance use: state of the art and future directions. Neuropsychobiology 2022;81(5):403-417. [doi: 10.1159/000527663] [Medline: 36349761]
41. Caponnetto P, Polosa R. Common predictors of smoking cessation in clinical practice. Respir Med 2008 Aug;102(8):1182-1192. [doi: 10.1016/j.rmed.2008.02.017] [Medline: 18586479]
42. Jarvis MJ, Sutherland G. Tobacco smoking. In: Bellack AS, Hersen M, editors. Comprehensive Clinical Psychology: Pergamon; 1998:645-674. [doi: 10.1016/B0080-4270(73)00101-2]
43. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]

Abbreviations

CoI: compatibility interval
CONSORT: Consolidated Standards of Reporting Trials
NEXit: Nicotine Exit
OR: odds ratio
RCT: randomized controlled trial

Edited by L Buis; submitted 24.Jun.2024; peer-reviewed by H Tateno, K Bold, MC Fahey, M Mansour, T Lid; revised version received 21.Nov.2025; accepted 28.Dec.2025; published 11.Feb.2026.

Please cite as:
Crawford J, Blomqvist J, Ulfsdotter Gunnarsson K, Bendtsen P, Bendtsen M
Individualized Treatment Effects of a Digital Smoking Cessation Intervention Among Individuals Looking Online for Help: Secondary Analysis of a Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e63578
URL: https://mhealth.jmir.org/2026/1/e63578
doi:10.2196/63578

Quality and Multifunctionality in Mobile Apps for Gestational Diabetes: Systematic App Review

Qimeng Zhao[1], MSc; Alison Cooke[1,2,3], PhD; Lishan Huang[1], MSc; Yimin Tang[1], MSc; Dawn Dowding[1], PhD

[1]Division of Nursing, Midwifery and Social Work, Faculty of Biology, Medicine and Health, University of Manchester, Room 6.306, Jean McFarlane Building, Oxford Road, Manchester, United Kingdom

[2]Centre for Research & Education Excellence (CeNREE), Royal Stoke University Hospital, University Hospitals of North Midlands NHS Trust, Stoke on Trent, United Kingdom

[3]School of Nursing and Midwifery, Keele University, Keele, United Kingdom

Corresponding Author:
Qimeng Zhao, MSc
Division of Nursing, Midwifery and Social Work, Faculty of Biology, Medicine and Health, University of Manchester, Room 6.306, Jean McFarlane Building, Oxford Road, Manchester, United Kingdom

Abstract

Background: The use of mobile health (mHealth) apps can assist with the management of gestational diabetes (GDM). Although a number of studies have demonstrated their efficacy in improving maternal-fetal outcomes, opinions differ regarding their usability and overall quality. Poorly designed apps, with ill-conceived features or inappropriate content, may pose a threat to patient safety. Nevertheless, very few studies provide in-depth evaluations of app design quality, and the diversity of features and techniques used remains insufficiently explored.

Objective: We aimed to evaluate the quality and multifunctionality of commercially available mHealth apps for GDM.

Methods: This is a systematic app review guided by the TECH (target user, evaluation focus, connectedness, and health domain) framework and the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist. Searches were conducted on the Apple App Store and Google Play. Apps were screened by name, description, and full navigation to identify inclusions. The quality of the apps was evaluated using the Mobile App Rating Scale and IMS Institute for Healthcare Informatics Functionality Score. Multifunctionality of the apps was evaluated using the GDM-adapted features and techniques list developed from the App Behavior Change Scale, NICE (National Institute for Health and Care Excellence) 2015 guidelines, and previous studies. The general features list, which contains instruction, data security, customization, and technical issues, was derived from previous studies.

Results: The search (June 2024) identified 23 commercially available apps from UK app stores. The overall app quality was evaluated to be satisfactory (Mobile App Rating Scale: mean 4.0, SD 0.36; IMS Institute for Healthcare Informatics Functionality Score: mean 5.83, SD 3.03). The multifunctionality evaluation found that the apps had a mean of 17.95 and SD of 7.31 across all 45 items. Overall, our findings suggested that mHealth apps for GDM achieved a certain level of multifunctionality. However, their feature types and supporting digital techniques are relatively basic. The apps focused on education and managing blood glucose control rather than integrating other self-monitoring data and pregnancy-relevant management into their design. The digital techniques used to achieve these features included text and manual operation, rather than other automated features.

Conclusions: This is the first app review to consider the relationship between app features and usability for women with GDM. Future app development should integrate a wide range of pregnancy-relevant information and more automated features and use advanced digital techniques to enable a holistic digital solution for women with GDM.

(JMIR Mhealth Uhealth 2026;14:e76862) doi:10.2196/76862

KEYWORDS

diabetes, gestational; pregnant woman; smartphone; mobile health; mHealth; app; behaviour change; mobile phone

Introduction

Challenges in Gestational Diabetes Self-Management

The self-management of gestational diabetes (GDM) usually includes a series of complex lifestyle behavioral changes such as diet, exercise, and continuous self-testing of blood glucose [1]. This systematic approach aims to help women maintain optimal blood glucose levels and minimize complications for both mother and baby. However, previous research highlights the challenges for women with GDM, who find it hard to understand or follow the instructions from doctors [2]. Women have to make decisions on food choices daily, and this is done through effective cooking strategies [3]. It also challenges

women to understand the labels and nutritional values of food [4]. Guidelines across countries emphasize the importance of strengthening physical activities and regard aerobic activities as acceptable approaches [5]. However, this requires carefully choosing the level of intensity and the types of exercise [5].

Mobile Health Apps Help With Self-Management of GDM

GDM digital technologies have covered various areas of GDM management, providing features including aspects of educational information, health behavior coaching, data recording, and communication interfaces for women [6]. By analyzing the overall effectiveness, previous literature found that GDM digital technologies could provide comparable quality of care to face-to-face visits with health care professionals or midwives, manifesting in heightened maternal health conditions and increased rates of natural childbirth, alongside diminished occurrences of adverse maternal and neonatal outcomes [7-9].

There are various forms of digital technologies, such as mobile apps, webpages, and digital devices, among which mobile apps are one of the most common types of technology used [9]. The rise in global smartphone penetration has led to the evolution of mobile apps, which are now one of the most representative forms of GDM digital technologies [8-10]. Mobile phones possess advantages in terms of accessibility as they are portable devices, thereby facilitating women's access to digital services irrespective of their geographical locations [11,12].

Variation in Perceived Usability of Mobile Health Apps for GDM

However, unlike the overall optimistic results from the evaluation of effectiveness, our previous review and other studies found a variation in pregnant women's views on the app's usability [13,14]. Women had positive views of apps when they supported self-monitoring and viewed them as beneficial tools for obtaining information about GDM, making informed dietary decisions, and facilitating exercise regimens [13,15-17]. The technologies also led to increased levels of adherence, concentration, and satisfaction among women [7,14]. The apps offered platforms for the women and clinicians to keep continuous interaction, fostering the client-clinician relationship [16,18]. This mitigated various concerns that women encounter in managing their condition, demonstrating efficacy in facilitating women's self-management practices [7-9]. In the meantime, apps added to women's emotional burden with problematic features and content [13]. For example, redundant processes and frequent failures in data recording reduced women's motivation for using the technology [18,19]. The differences between the information provided by apps and clinicians confused the women [16]. Evidence suggests that mobile health (mHealth) apps might unintentionally lead to increased unhealthy behavior [20]. mHealth apps designed inappropriately may contain inaccurate or incomplete information that misleads users or provide insufficient support for managing high-risk behaviors, which could affect users' safety [21].

Challenges to Conducting a Comprehensive Evaluation of GDM Apps

To ensure that women are offered apps of the best quality, health care practices face challenges in evaluating and selecting suitable options. The current era is facing the results of unregulated mHealth markets, which flourish in the number of emerging mHealth apps but vary in quality [22]. The increasing study of app quality evaluation has triggered a discussion on the definition of app quality. Studies found that the criteria used for evaluating app quality were heterogeneous [23,24]. In addition, established criteria lack precise definitions, which hinders the assessment of app quality.

Two previous reviews have evaluated the functionality and usability of GDM mHealth apps [6,25]. Both reviews primarily relied on a single quality criterion: the Mobile App Rating Scale (MARS). However, neither study addressed how well app features are adapted to GDM self-management content. Similarly, they offered no practical insights to assist health care professionals in selecting apps that are both functional and user-friendly. What is missing from the current literature is an evaluation of how these apps contain features suitable for GDM care, and guidance that helps health care professionals when selecting usable apps.

Objectives

This study reviewed commercially available mobile apps to explore app quality and multifunctionality, targeted at antenatal care for women with GDM or pre-existing diabetes during pregnancy, available in the UK app market, and in the English language. This systematic evaluation only includes free-download apps and analyzes the available free-of-charge features, providing insights into the types of features that are most accessible for women with GDM. We assessed the quality of the apps using standardized evaluation scales. We evaluated the app's multifunctionality based on app features and content tailored specifically to the GDM population, as well as general features that affect the normal use of apps.

Methods

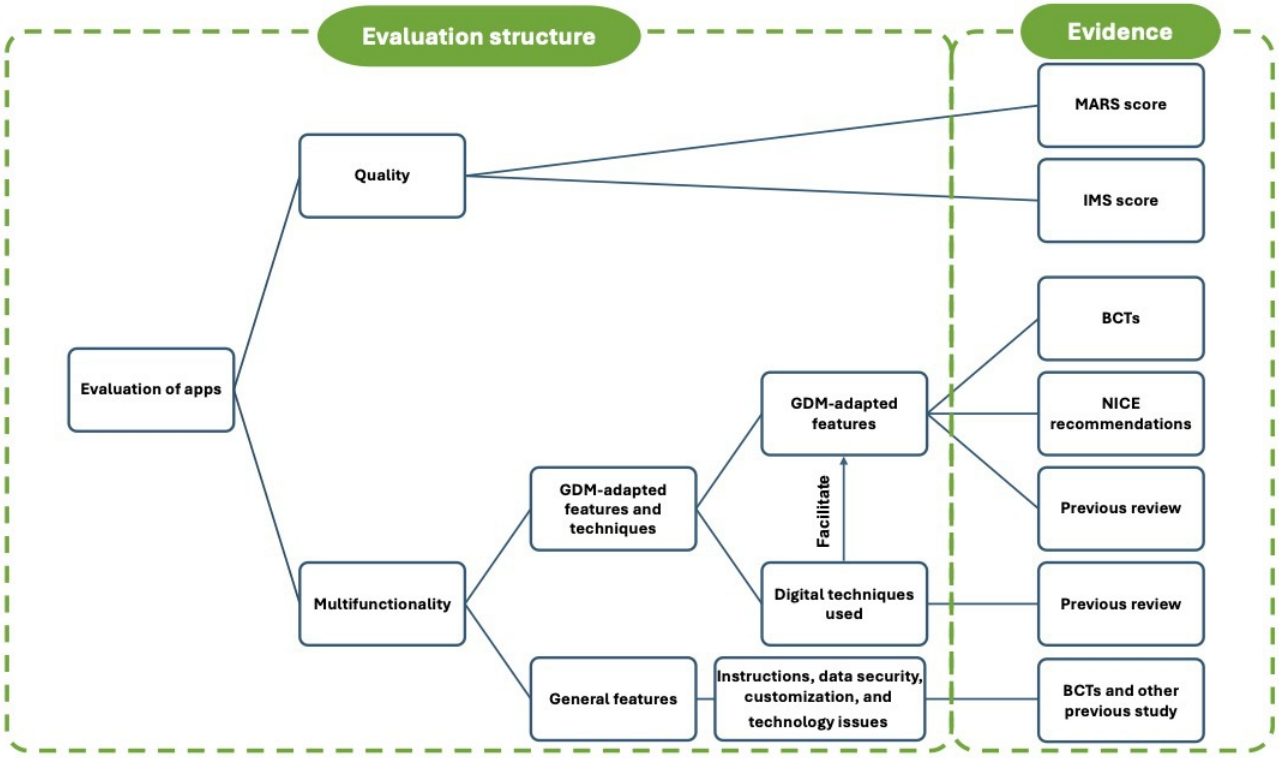
Study Design

This study followed the TECH (target user, evaluation focus, connectedness, and health domain) framework proposed by Gasteiger et al [26] (2023), which is tailored for the context of reviewing and evaluating the quality of commercially available mHealth apps as opposed to the systematic literature review methodology. Informed by the traditional systematic review process, the TECH framework aims to help conduct reviews on mobile apps by establishing search strategies, the use of eligibility criteria, and a comprehensive quality appraisal. It provides guidance to conduct a search, and has been applied in recent studies to evaluate mHealth apps, including risk assessment and self-management tools, by examining both their functionality and content [27-29].

This study consists of 2 major domains of evaluation: app quality and app multifunctionality. Figure 1 shows the structure of the evaluation elements and the evidence supporting the evaluation. The app quality evaluation in this study aimed to

explore the type of features being used in the included apps and used generalized tools to quantify the level of quality. The measurement was informed by the TECH framework and guidance by introducing both the MARS and the IMS Institute for Healthcare Informatics Functionality Score (IMS). The aggregation of the 2 quality appraisal tools helps to evaluate mobile apps more comprehensively.

Figure 1. Evaluation structure. BCT: behavior change technique; GDM: gestational diabetes; IMS: The IMS Institute for Healthcare Informatics Functionality Score; NICE: National Institute for Health and Care Excellence; MARS: Mobile App Rating Scale.



The multifunctionality evaluation aimed to conduct a GDM context-tailored assessment on the behavior change techniques (BCT) used in the apps, as well as other essential features needed to be enabled for normal app usage. Therefore, there were two elements in the multifunctionality evaluation: (1) the GDM-adapted features and techniques, which refer to the app features aligned with BCTs, along with the digital techniques used to enable the features; and (2) the general features, which refers to the essential features that enable the normal usage of apps, including instructions, data security, customization, and technical issues.

Eligibility Criteria

Following the guidance of the TECH framework, a scoping search on Apple App Store and Google Play in the UK context was run in March 2024 to identify eligibility criteria [26]. This study aims to identify apps that include GDM as one of the targeted user groups and that eligible apps must refer to GDM in their description to qualify for inclusion in this study. From this process, three potential types of mobile apps were identified: (1) mobile apps for GDM management, (2) mobile apps for pregnancy management with additional features for blood glucose control, and (3) mobile apps tailored for diabetes patients but claimed to be suitable for women with GDM. Additionally, the mobile apps were available in the UK context and the English language, free to download, running on Android (Google LLC) or iOS (Apple Inc). The eligibility criteria are listed as follows (Multimedia Appendix 1):

1. Target user (T): mobile apps tailored for GDM, pregnancy, and diabetes users, with GDM as one of the targeted user groups.
2. Evaluation focus (E): content of information, quality, and functionality of the mobile apps that are suitable for users with GDM.
3. Connectedness (C): mobile apps with external devices such as glucose meters and smartwatches.
4. Health domain (H): mobile apps suitable for antenatal behavioral change for women with GDM rather than prevention or postpartum care.

Search

The search keywords "gestational diabetes," "GDM," and "pregnancy diabetes" were applied in the search strategy. The final search was conducted by 3 reviewers on the Apple App Store and Google Play in June 2024. Due to the Apple App Store's online page design, it was not possible to export the search results into screening software to remove duplicates. Additionally, the mobile app search results were subject to frequent changes over time. To avoid unnecessary duplication in the search, all 3 reviewers searched and performed screening on the same day.

Screening

The screening process was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist and the TECH methodology [26,30]. Two steps of

screening were applied, including title screening and full mobile app screening. The mobile app name and description screening process was conducted on the search results pages to identify potential mobile apps for inclusion. We contacted the app companies to access those apps with limited access during the retrieval process. A subsequent full mobile app screening was conducted to check the relevance and feature operability. To include the mobile apps which were usable at the technological level as well as suitable for the targeted population, the full mobile app screening process made a decision of inclusion when (1) the screened mobile app was available in English; (2) no constant technical glitch or crash occurred when operating; and/or (3) the mobile app mentioned GDM at least once in options, information, or features. A snowballing strategy was used after the full mobile app screening process to check the relevant-recommended apps generated within the search results pages of the final-included mobile apps, such as the "you might also like" and "similar apps" pages, to reduce missing mobile apps.

During the screening process, 2 pairs of authors independently screened the search results in Google Play and Apple App Store. The level of agreement between the reviewers was determined using Cohen κ to calculate interrater reliability. To minimize the potential for duplication inherent in app store searches, the analysis was based on the results from a single search term. Discrepancies between the paired authors were discussed and resolved. The results of eligible mobile apps were checked and agreed upon by the other authors. The devices used to search and screen the apps were iPhone 11 Pro Max (iOS 17.5), iPhone 11 (iOS 17.5), iPhone X (iOS 16.7.8), Samsung Galaxy J5 (Android v8.1 Oreo), and Huawei Mate 10 (HarmonyOS 4.0).

Quality Appraisal

Overview

Two quality appraisal tools, which are widely applied for mHealth evaluation, the MARS [26,31] and the IMS tool [32], were used to evaluate the mHealth apps. The MARS (26 items) was used to evaluate the technical design, and the IMS (11 items) to evaluate the number of functions.

The MARS was scored from inadequate to excellent with a corresponding score from 1 to 5 in the 5-point Likert scale. This review followed the recommendation of MARS to calculate MARS scores using the 16 items of app quality rating. The subjective quality and perceived impact items were combined as subjective quality (10 items) for evaluation in this review. The IMS [32] items were coded as 1 per item if app features were presented and otherwise 0 per item, generating a functionality score ranging from 0 to 11 for each app. The overall mapping scores representing the frequencies of applied features were presented in a radar graph as recommended [32].

Quality appraisal was conducted by 2 pairs of reviewers. The agreement rate between the reviewers using different scales was calculated, with interclass correlation coefficients (ICCs) for MARS results [33] and Cohen κ for the IMS results. The level of agreement in ICCs, which ranges between 0 (no agreement) and 1 (perfect agreement), was reported in the value and followed the level categorization as either poor (<0.40), fair (0.40 - 0.59), good (0.60 - 0.74), or excellent (0.75 - 1.0) [34]. The calculation was conducted using IBM SPSS (version 29; IBM Corp).

Data Extraction

Data extraction focused on two categories: (1) app characteristics, which included basic information such as the app operating system, version, and size; and (2) multifunctionality, which included GDM-adapted features and techniques, as well as general features. A list of uncategorized app features and techniques was initially extracted and used to inform the GDM-adapted features and techniques list and the general features list. The two lists contributed to the evaluation of the app's multifunctionality (Multimedia Appendix 2). After the development of the evaluation lists, the apps were reviewed again by 1 author and extracted for the items on the lists. The extracted features and techniques of individual apps were checked and discussed to reach an agreement in the author group.

The GDM-adapted features and techniques lists included types of app features (GDM-adapted features) and digital techniques (GDM-adapted techniques). GDM-adapted features refer to the app feature content, which served as a digital solution to clinical interventions for GDM management. For example, blood glucose data recording served as a digital solution to a paper diary for blood glucose documentation. GDM-adapted techniques are the digital techniques used to enable the app features. For example, Bluetooth (Bluetooth Special Interest Group) techniques and manual recording are both digital techniques used to enable the data recording app feature. General features referred to instructions, data security, and customization, which were extracted based on other studies [35,36].

The development of the GDM-adapted features and techniques list included app features aligned with BCTs [36], NICE (National Institute for Health and Care Excellence) 2015 recommendations [37], and our previous review on women's preferences [13]. BCTs include strategies such as education, goal setting, self-monitoring, feedback, and reinforcement that encourage healthy behaviors [36]. The 2015 NICE guideline recommends managing GDM through individualized diet and exercise, along with blood glucose monitoring [37]. It also includes pregnancy management guidance by including frequent antenatal visits and monitoring of fetal growth and well-being [37]. Our previous review found that women with GDM valued digital tools with education, personalized guidance, and easy data management [13]. It highlights the need for more usable technologies that support self-monitoring of GDM [13]. Integrating these sources created a clear framework for assessing app features, making sure the GDM-adapted features and techniques list reflects behavior change, clinical care, and women's needs when managing GDM. Additionally, some basic digital techniques, such as text-based educational information and data recording approaches, were added to the final list of app features without direct indications from literature or recommendations.

Data Analysis and Presentation

The extracted features and techniques of apps were analyzed and presented in aggregated data. Categorical data, including app characteristics items, GDM-adapted features, GDM-adapted techniques, and general features, were summarized as frequencies and percentages. Continuous data, including app size, MARS score, and IMS score, were presented as means with SDs. App characteristics, MARS score, GDM-adapted features and techniques, and general features were summarized in tables to provide a clear presentation of results. The IMS score was presented in a radar chart to illustrate the distribution of IMS functionalities across 10 domains.

Ethical Considerations

This study did not involve human participants, animals, or any collection of personally identifiable information. All data analyzed were publicly available in the Apple App Store and Google Play Store. Accordingly, institutional review board approval was not required. As no human subjects were enrolled, informed consent was not applicable.

Results

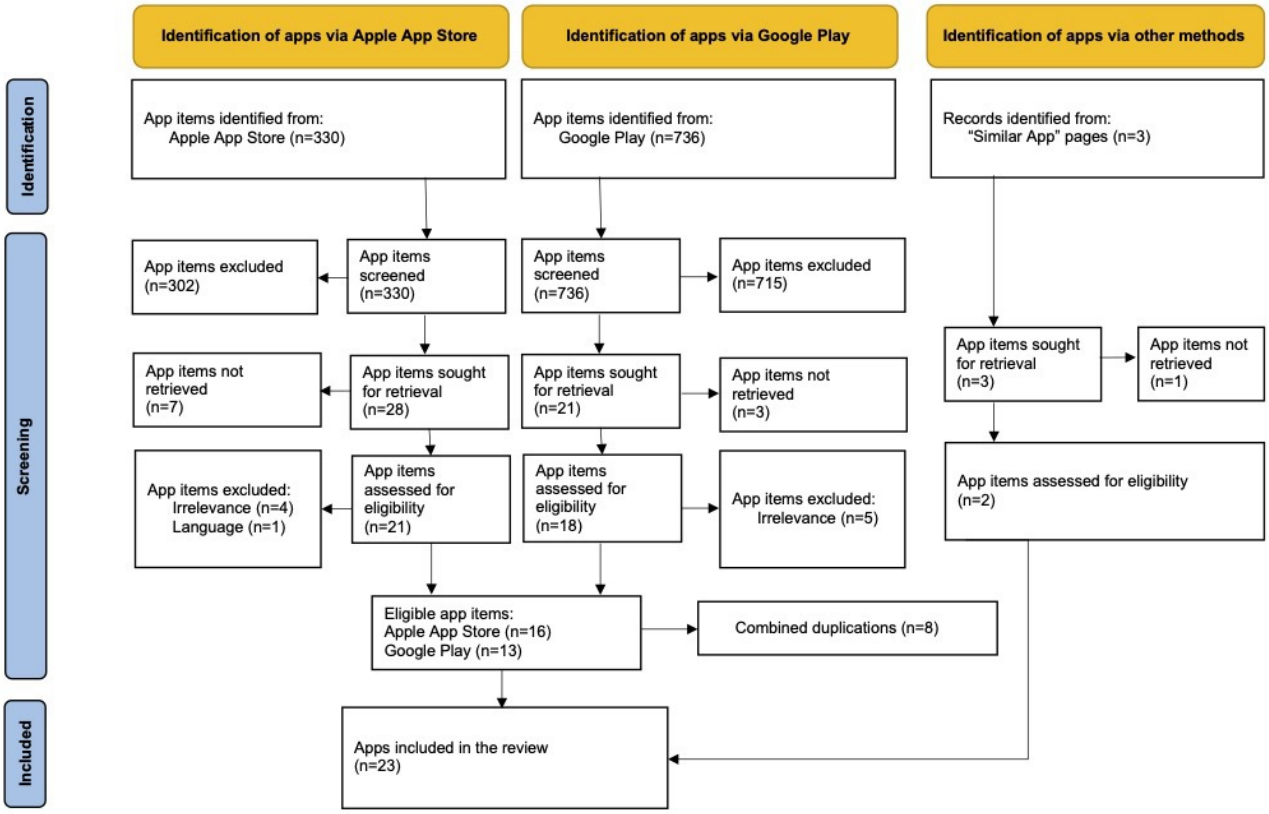
Search Results

The search identified 1066 apps, 330 from the Apple App Store and 736 from Google Play. Two paired authors independently screened the search results in both app stores, the Apple App Store, resulting in 28 eligible mobile apps from the Apple App Store, and 21 in Google Play. The interrater reliability in both pairs indicates a minimum of substantial agreement, with Cohen κ values of 0.62 and 0.91. The title and description screening process resulted in 49 potentially eligible mobile apps in the two app stores. The mobile apps available in both operating systems were not combined at this step.

The 49 eligible mobile apps were retrieved and installed before the full mobile app screening process. This process led to 10 exclusions due to activation (n=7), subscription (n=1), and technical crash issues (n=2). The retrieving, installing, and checking process led to 39 (n=21 in iOS, n=18 in Android) inclusions for the full mobile app screening. The full app screening led to 10 exclusions (n=5 in iOS, n=5 in Android). The reasons for exclusion included irrelevance (n=9) and language (n=1). Perfect agreement was identified between the 2 pairs of authors. Mobile apps available in both operating systems (n=8) were then combined in number, resulting in 21 eligible mobile apps.

A subsequent snowballing process generated the inclusion of 2 mobile apps (available in both operating systems). The search and screening process contributed to the eventual inclusion of 23 eligible mobile apps for this study. Figure 2 illustrates the PRISMA diagram for the screening process (Multimedia Appendix 3 for the enlarged PRISMA diagram).

Figure 2. PRISMA diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



App Characteristics

Of the 23 reviewed apps, 56% (n=13) were adapted to both operating systems (iOS and Android), followed by 22% (n=5) of the apps which were exclusively for iOS, and 22% (n=5) of the apps which were exclusively for Android. Most (n=16, 70%) of the apps were updated in 2024. On average, the apps were of mean 66.43 and SD 53.46 MB, ranging between 6.2 and

223.4 MB. All the apps were free to download, with 48% (n=11) free to access all features, and 52% (n=12) required a subscription. The major targeted user groups of the apps were people with diabetes (n=12, 52%), women with GDM (n=7, 30%), followed by women who are pregnant (n=3, 13%) and women with diabetes who are pregnant (n=1, 4%).

We categorized the apps into 4 major types based on the characteristics of features: multipurpose, self-monitoring, educational, and communicational. Self-monitoring apps were identified by their ability to assist women's behavior for self-monitoring, such as recording data and reminding them of self-tests, with a total of 5 apps. Educational apps were identified by their ability to offer women information, with a total of 5 apps. Communicational apps were identified by their ability to provide a platform for women to communicate with their peers and health care professionals, with a total of 2 apps. The apps that used more than one of the mentioned characteristics of features were identified as multipurpose apps (n=11). All apps (n=23) were free to download, with all apps that were tailored for GDM or pregnant users (n=11) being free to access all features; meanwhile, the other 12 apps for general diabetes users needed to subscribe for full access. Table 1 shows the characteristics of the reviewed apps.

Table . App characteristics.

Evaluations	Values
Operating systems, n (%)	
Both	13 (56)
iOS	5 (22)
Android	5 (22)
Last update, n (%)	
2024	16 (70)
2023	4 (17)
2022 and before	3 (13)
Size (MB), n (%)	
0 - 50	11 (48)
50 - 100	7 (30)
>100	5 (22)
Targeted population, n (%)	
Diabetes	12 (52)
GDM[a]	7 (30)
Pregnancy	3 (13)
Diabetes in pregnancy	1 (4)
App type, n (%)	
Multipurpose	11 (48)
Educational	5 (22)
Self-monitoring	5 (22)
Communicational	2 (9)
Price, n (%)	
Subscription for full access	12 (52)
Free full access	11 (48)
MARS[b] quality score, mean (SD)	4.0 (0.37)
IMS[c] score, mean (SD)	5.83 (3.10)

[a]GDM: gestational diabetes.

[b]MARS: Mobile App Rating Scale.

[c]IMS: The IMS Institute for Healthcare Informatics Functionality Score.

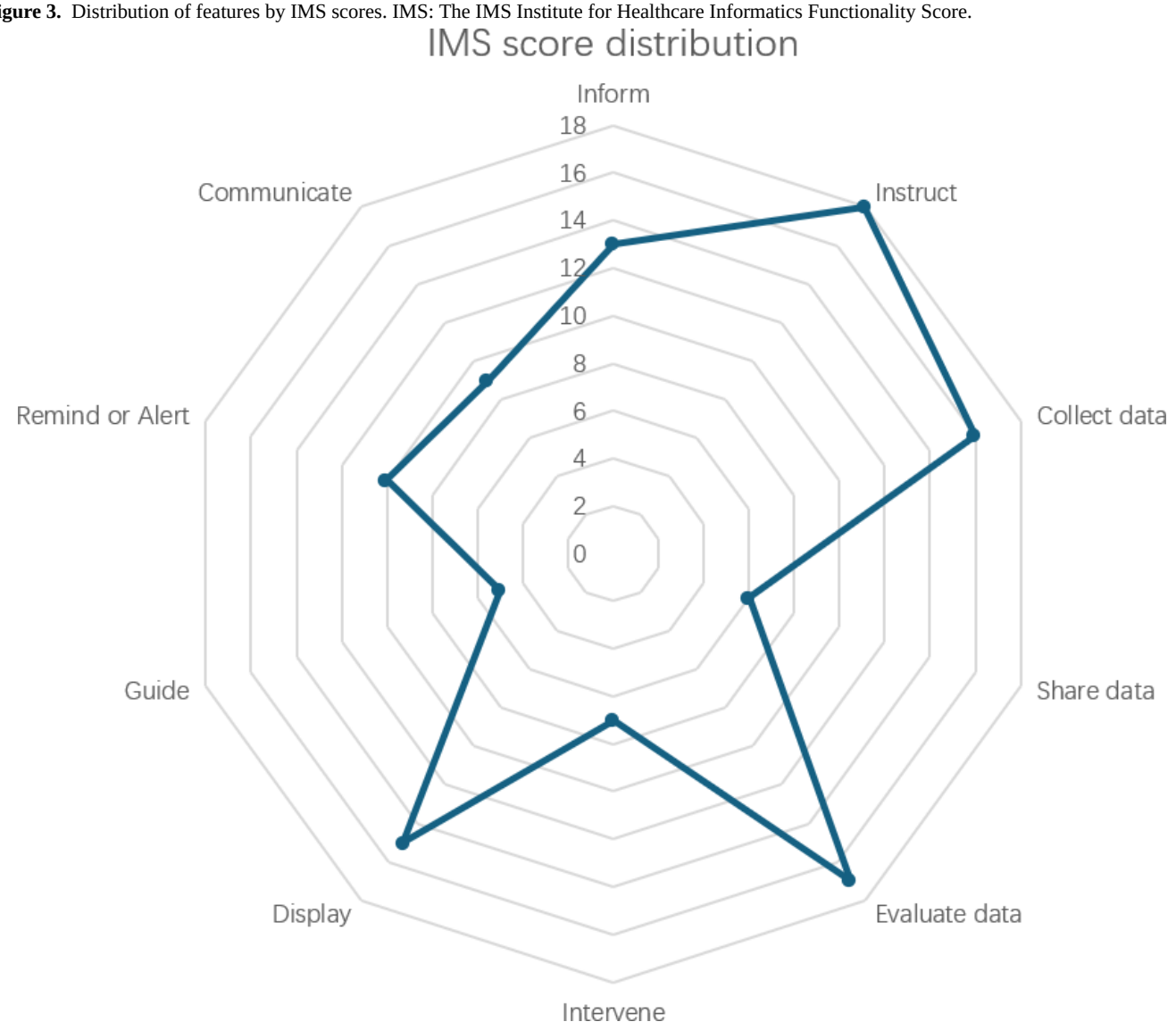
Quality Appraisal

Overview

Overall, the mean MARS score was 4.0 and SD 0.36, and the mean IMS score was 5.83 and SD 3.03 (see details in Multimedia Appendix 4). The mean MARS quality score indicates a satisfactory level of quality, which appears to be higher than the average MARS quality score reported for mHealth apps in the literature (mean 4.0, SD 0.36 vs mean 3.51, SD 0.71) [38]. Regarding individual MARS quality score domains, the MARS engagement domain, which is relevant to coaching and personalization features, scored the lowest (mean 3.56, SD 0.60) compared to other MARS domains (MARS functionality: mean 4.48, SD 0.72, MARS esthetics: mean 4.12, SD 0.48, MARS information: mean 4.02, SD 0.44). Apps were also evaluated based on their different types of functionality using the MARS quality score: the multipurpose apps, educational apps, self-monitoring apps, and communicational apps. Multipurpose (mean 4.06, SD 0.44) and self-monitoring (mean 4.02, SD 0.25) apps appeared to score higher than other types of apps (educational apps: mean 3.92, SD 0.37, communicational apps: mean 3.84, SD 0.47).

The mean IMS score indicated a moderate level of the number of app features. Multipurpose (mean 7.91, SD 2.12) and self-monitoring (mean 6.60, SD 1.14) apps tended to score higher than the other two types of apps. Figure 3 shows the radar graph of the distribution of features scored by the IMS. Among all the 10 IMS features, instructions (n=18), data management features (n=18), and information (n=13) were the predominant features used by the 23 reviewed apps. The guide feature scored lowest compared to other IMS features (n=5).

Figure 3. Distribution of features by IMS scores. IMS: The IMS Institute for Healthcare Informatics Functionality Score.



Reviewers had substantial agreement on their evaluation of iOS apps; IMS (κ=0.834, 95% CI 0.74-0.93, *P*<.001) and MARS (ICC=0.808, 95% CI 0.742 - 0.858) scores. A similar level of agreement was observed for Android apps, with IMS (κ=0.851, 95% CI 0.76-0.95, *P*<.001) and MARS (ICC=0.755, 95% CI 0.67 - 0.82).

Multifunctionality

The analysis of features and techniques was based on the exploration of features that would need to be integrated into an app to assist specifically with GDM. We incorporated the BCTs [36], NICE recommendations [37], and our previous review on women's preferred features [13]. The process generated a list

of GDM-adapted features and techniques (Table 2, details in Multimedia Appendix 5). The features were categorized into 5 domains: education, data management, coaching features, communication features, and pregnancy management.

Table . GDM[a]-adapted features and techniques (N=23).

GDM-adapted features and techniques	Evidence[b]	Frequencies of applied features
Education (n=18)		
Educational content features		
Knowledge about GDM diagnosis	N, P	10
Treatment types for GDM	N	7
Low GI[c] index food choices	B, N	15
Food recipes	B	4
Evidence-based information	B, P	11
Self-monitoring skill	B, N, P	8
Educational information delivery techniques		
Text	__[d]	18
Pictures or graphs	P	9
Videos	P	4
Data management (n=16)		
Data management content features		
Blood glucose	N, P	14
Food	N, P	8
Exercise	N, P	7
Abnormal symptom reporting	N	4
Medication	N	6
Data recording techniques		
Manual (number or text)	—	14
Automatic glucose data transmission	P	4
Automatic exercise data transmission	P	3
Food or exercise database	P	7
Image recognition techniques	P	3
Data visualization techniques		
Tables with text	—	11
Graphs	P	15
Combination charts	P	3
Other data management features		
Digital feedback	B, N, P	14
Data export	B	8
Personal-tailored information or suggestions	B, N, P	3
Coaching features (n=11)		
Reminders	B, P	9
Compliance motivator	B, P	5
Rewarding for milestone achievements	B	3
Communication features (n=9)		
Communication with HCPs[e]	B, N, P	6
Communication with peers	B, P	6
Pregnancy management features (n=5)		

GDM-adapted features and techniques	Evidence[b]	Frequencies of applied features
Pregnancy progression education and management	N, P	4
Fetal monitoring	N, P	3

[a]GDM: gestational diabetes.

[b]Evidence: B: behavior change techniques; N: NICE (National Institute for Health and Care Excellence) recommendations; P: women's preferred features.

[c]GI: glycemic index.

[d]Not applicable.

[e]HCP: health care professional.

Table 3 summarizes the predominant GDM-adapted features and techniques applied by the reviewed apps. The analysis reveals a strong emphasis on educational content (n=18) and data management (n=16) within the mobile apps. The educational components of the apps focus on delivering content critical for GDM management, mainly including knowledge about GDM diagnosis (n=10) and low GI index food choices (n=15). Data management predominantly focuses on collecting blood glucose data, with over 50% (n=14) of apps, rather than food (n=8) and exercise tracking (n=7).

Table . Predominant applied GDM[a]-adapted features and techniques.

App names	Predominant educational features		Predominant data management features	Predominant techniques		
	GDM diagnosis	Food choices	Blood glucose	Text information delivery	Manual data recording	Text data visualization
GDm-Health	✓	✓	✓	✓	✓	✓
Malama	✓	✓	✓	✓	✓	✓
MyManis	✓	✓	✓	✓	✓	✓
Flora	✓	✓		✓		
hug + u			✓	✓	✓	✓
myFetalLife			✓	✓		
Blood Sugar - Diabetes App		✓	✓	✓	✓	✓
Blood Sugar Tracker - Diabetes	✓		✓	✓	✓	✓
Diabetes:M	✓	✓	✓	✓	✓	✓
DiabTrend		✓	✓	✓	✓	
MyNetDiary			✓	✓	✓	✓
Gestational Diabetes Diet		✓		✓		
glukoGest	✓			✓		
Pregnant with diabetes	✓	✓		✓		
NewToType2	✓	✓		✓		
Pingoo	✓	✓		✓		
Gestational Diabetes Tracker			✓		✓	✓
Carbs & Cals		✓				✓
Diabetic diary-Glucose tracker			✓		✓	✓
forDiabetes			✓		✓	✓
mySugr			✓		✓	
Hlgedi		✓		✓		
Diabetes Forum		✓		✓		

[a]GDM: gestational diabetes.

Text-based educational information was nearly ubiquitous, and core tracking features—particularly for blood glucose—were widely implemented. Text-based information delivery was the most common mode, used in over 50% (n=18) of apps. Manual data recording was predominant (used in 14 apps), and 12 used tables (text) to visualize the data. Among all 23 apps, Malama (a multifunctional, GDM-focused app) and Diabetes:M (a multifunctional, diabetes-focused app) cover all the predominant features and techniques.

We identified the features and techniques applied by fewer than 4 apps as the least applied features. The least addressed features included the provision of food recipes (n=4), abnormal symptom reporting (n=4), rewarding for milestone achievements (n=3), and pregnancy progression and education management (n=4). The least addressed techniques included video information delivery (n=4), automatic blood glucose data transmission (n=4), automatic exercise data transmission (n=3), image recognition techniques (n=3), combination charts (n=3), and personal-tailored information or suggestions (n=3).

General Features

Overall, most reviewed apps used general features which ensured normal app usage, with 23 apps addressing data security, 22 apps addressing instructions, and 19 apps addressing personalization. There were 5 apps identified as having technical issues that affect normal use, or distractions that disrupt user interactions. General features and technical issues are summarized in Multimedia Appendix 6.

Some general features were applied universally by the apps. These include consenting to privacy policies (n=23) and asking permission to access data (n=23), providing a quick step-by-step introduction (n=16) when users launch the apps. For personalization, collecting users' demographic data (n=16), asking their preferences for measurement units (n=14), and enabling customized features (n=16), such as reminder timing, were used. Very few apps mentioned the location of data storage (n=5) and the techniques (n=4) used to protect user data.

Additionally, of the 23 evaluated apps, there were 7 with issues that could affect normal user experience, including 5 multipurpose apps (MyManis, hug + u, myFetalLife, Blood Sugar - Diabetes App, and Diabetes:M) and 2 self-monitoring apps (Carbs & Cals and Diabetic diary-Glucose tracker). There were three major issues among these apps: (1) frequent technical glitches, (2) data management issues, and (3) frequent advertisements. Data management issues existed in 4 multipurpose apps, with frequent failure or errors of data recording (MyManis and Diabetes:M), limitations in the data recording feature when inputting decimals and time points (hug + u), and failure of displaying the recorded data in 1 app (myFetalLife). Frequent advertisements took place in 3 apps (Blood Sugar - Diabetes App, Carbs & Cals, and Diabetic diary-Glucose tracker), with advertisements showing up frequently when moving between screens or inputting data, leading to a sense of distraction. Technical glitches and crashes happened to 1 multipurpose app (myFetalLife). While this app contained various features suitable for the management of GDM, technical issues took place frequently and caused crashes and failures when recording the data.

Usable Apps

This review evaluated the app functionality based on the generated lists of GDM-adapted features and techniques, as well as general functionality, collectively with 45 evaluation items on features and content. On average, the apps applied a mean of 17.95 and SD of 7.31 of all 45 items across domains. For GDM-adapted features and techniques, the apps applied a mean of 10.82 and SD of 5.31 of the 32 items. For general functionality evaluation, the apps applied a mean of 7.13 and SD of 2.58 of the 13 items.

We considered apps usable when they contained features and content higher than the average value in each domain and the total average value, with no distractions or technical issues (Table 4). These 6 apps were either categorized in the multipurpose (n=4) or the self-monitoring (n=2) app type in this review.

Table . Usable apps and evaluation.

	App type	Health care domain	MARS[a] quality	IMS[b] score	Multifunctionality and content	General functionality
Full score or number of features	N/A[c]	N/A	5	11	32	13
Average score or addressed features, mean (SD)	N/A	N/A	4.0 (0.36)	5.83 (3.03)	10.82 (5.31)	7.13 (2.58)
Malama	Multipurpose	GDM[d]	4.63	9	24	9
GDm-Health	Multipurpose	GDM	4.47	8	13	11
MyNetDiary	Multipurpose	Diabetes	4.40	10	16	12
DiabTrend	Multipurpose	Diabetes	4.03	9	20	9
MySugr	SMBG[e] app	Diabetes	4.13	8	14	11
forDiabetes	SMBG app	Diabetes	4.3	6	11	9

[a]MARS: Mobile App Rating Scale.

[b]IMS: The IMS Institute for Healthcare Informatics Functionality Score.

[c]N/A: not applicable.

[d]GDM: gestational diabetes.

[e]SMBG: self-monitoring of blood glucose.

Discussion

Overview

This study aimed to provide a comprehensive evaluation of the quality and multifunctionality of commercially available apps on their cash-free features for GDM self-management. By evaluating app quality using the MARS quality and IMS scales, the results showed that the overall app quality is at a satisfactory level. However, regarding individual MARS domains, the MARS engagement domain, which contains coaching and personalized content, was scored lowest among other domains. Following this, the multifunctionality of apps was evaluated by two elements: (1) GDM-adapted features and techniques, namely the app features that fitted into GDM management context, and the corresponding digital techniques used to enable the features, explored the variety of features used by the reviewed apps; (2) general features, which included instruction, data security, customization, and technical issues, explored the app features regarded as essential for basic app usage.

Primary Findings: Basic Features, Limited Content, and the Need for Advanced Design

The reviewed apps predominantly focused on GDM educational content and blood glucose data recording. In addition, the digital techniques used to enable such features were relatively basic. Most apps relied on text-based approaches for content presentation and required manual operation for data management. Moreover, their content reflected only the most fundamental requirement for GDM, focusing primarily on blood glucose management rather than addressing the broader pregnancy experience.

One of our primary findings is the narrow focus on blood glucose monitoring in the design of many apps. Many of the apps included in this review offered features that could support blood glucose self-management, such as educational resources on self-test skill education and dietary guidance, as well as data recording, digital feedback, and graphic visualization. These features were also identified in our previous study as being preferred by women with GDM, as they facilitated self-management [13]. For example, having educational information in a single app improved easy access and helped women recall information [13]. Initial digital feedback using colored labels, combined with graphic visualization, helped the women interpret their data and adjust their diet according to blood glucose trends [13]. However, most apps lacked a multidimensional approach to GDM management. Even though diet and physical activity are both essential for stabilizing blood glucose levels, very few apps in this review included relevant content and features.

Our findings suggest that features relevant to pregnancy progression management were lacking in available apps. In this study, only a small number of apps provided information or features to support pregnancy management and fetal monitoring. There were only 5 apps that included features or information relevant to pregnancy, such as informing about baby growth over time. The features and information about pregnancy and fetal well-being were more likely to be available in those apps designed for pregnancy care. However, the functionality of these pregnancy-tailored apps in blood sugar control was relatively underdeveloped, typically due to technical issues related to data recording and display. These findings indicated an imbalance in focus in app design, with limited integration of broader pregnancy-relevant needs.

In addition, automated features, including automated data transmission and app-generated behavior change suggestions, were used by only a few apps. Similar patterns are also observed in other mHealth apps for pregnancy care, where basic features and techniques still dominate the feature composition [39-41]. These apps tended to focus on either collecting users' personal and physiological data or providing normative and comprehensive information, rather than proactively facilitating women's behavior changes [39-41]. Automated features, such as rewarding on milestone achievements, identifying barriers for behavior change, or providing feedback for behavioral adjustment, are rarely incorporated into the design of pregnancy-related apps [39,40]. Instead, these tasks are sometimes managed through interactions with health care professionals and handled manually [41].

This indicates that the apps currently available and free to download appear to lag behind the evolving needs of women and health care professionals. Women seek tailored guidance for managing their blood glucose, which enhances women's self-awareness and supports their autonomy in self-management [13]. This is usually achieved through effective communication with health care professionals, where the women can receive one-to-one advice and personalized feedback on their blood glucose levels, diet, and lifestyle adjustment [42,43]. However, limited opportunities for antenatal consultations and the heavy workload of health care professionals have often resulted in limited advice on continuous behavior changes [42,43].

Implications for Designing Guidelines and Recommendations

Our findings highlight a gap between the potential of technology and its real-world application. Frontier automated techniques are advancing at an astonishing pace, such as deep learning, decision trees, and reinforcement learning, which have been explored for predicting blood glucose levels and providing personalized treatment recommendations [44,45]. These tools can continuously process data such as blood glucose levels, ketonuria, and dietary adherence, identifying patterns and making real-time adjustments [44]. However, our review revealed that many frontier technologies remain largely inaccessible to women with GDM.

To address this gap, it is essential to integrate broader pregnancy progression management into a condition-specific management app that requires thoughtful design to ensure relevance across different stages of pregnancy and to balance general maternal health management with condition-focused support. As recommended by the 2015 NICE guidance, pregnancy progression management is an important component of GDM management [37]. This requires additional attention on fetal growth and well-being, indicating that additional screening test results and fetal self-observation data recording should be considered [37]. Additional evidence found that women with GDM preferred mHealth technologies that were closely aligned with their pregnancy needs, such as reminders for antenatal visits and pregnancy stages, and they particularly valued educational information related to their pregnancies [16,46].

The findings emphasize the potential of digital tools to provide support for individualized care. It calls for greater integration of automated technologies into app design. This requires co-design approaches involving women and health care professionals to ensure that future apps move from basic features to personalized, interactive, and context-based support for GDM management. Additionally, further exploration is needed to evaluate the automated app feature designs regarding their usability, effectiveness, and safety in real-world contexts.

Recommendations for Development

Given the wide variation in features and techniques available in current GDM apps, and the range of factors that may influence app development, we adopted an evidence-based approach to summarize broadly applicable recommendations

for the development of apps for women with GDM (Textbox 1). These recommendations outline options for core functional design, while also identifying optional advanced features that may be incorporated where appropriate.

Textbox 1. App development recommendations.

Education

Information about gestational diabetes (GDM):

- Include information on the mechanism of GDM, its relevant risks, and impact on pregnancy. The information should aim to help women recognize the importance of monitoring their blood glucose levels.

Dietary management information:

- The rationale for dietary management and basic dietary strategies should be explained, with links to practical and reliable resources, such as guidance on low-glycemic index food choices.

Physical activity information:

- Provide information about the appropriate types of exercise for pregnant women, optimal timing for exercise, the importance of physical activity, and considerations for ensuring safety.
- Information should be provided on appropriate exercise duration and intensity, conditions in which exercise is not suitable, and when exercise should be stopped if certain symptoms occur.

Self-monitoring skills:

- Provide information about how to perform self-monitoring, including capillary blood glucose testing skills, appropriate timing of testing (for example, fasting, before or after meals, and before bedtime), and evidence-based target ranges for blood glucose levels.

Pregnancy-relevant information:

- Provide information on pregnancy progression, including the different stages of pregnancy and corresponding information about the baby, such as fetal growth.

Basic techniques:

- Text information delivery.

Advanced techniques:

- Graphs, pictures, and videos in information delivery.

Development considerations:

- The primary aim of providing information is to increase women's awareness of self-monitoring and, in turn, improve their skills. Information should be clear, easy to understand, and sufficient to support understanding.
- Aside from basic information, consider practical resources, such as food recipes or exercise tutorial videos, that support behavior changes.
- Resources provided should be evidence-based and reliable.

Data management

Types of data:

- Ensure blood glucose data is accurately recorded and clearly visualized in the app. Consider other essential data, including dietary intake, physical activity, symptoms, medication usage, fetal growth, and fetal well-being, to be recorded within the app.

Basic techniques:

- Manual data recording using text, data visualization using tables, and initial data evaluation by preset normal values.

Advanced techniques:

- Automated data transmission (eg, Bluetooth and image recognition techniques), graphic data visualization to show trends, combination charts to cross-check different types of data, and personalized suggestions based on recorded data.

Development considerations:

- Enable easy data recording, prioritizing automated digital techniques over manual text entry, and provide appropriate indicators and options to improve efficiency.
- Considering digital data analysis techniques to evaluate patients' recorded data, and where appropriate, to offer initial suggestions to support patients.

Communication

Features:

- Communication with health care professionals and peers.

Basic techniques:

- Easy-access contact buttons to start phone calls or send text messages.

Advanced techniques:

- Built-in chat box, online forum.

Development considerations:

- Ensure access to health care professionals within the app.
- The features should be designed to help address the women's questions in real time.
- Specific staffing roles should be considered to support and manage communication modules within the app.

Coaching, customization, and technical issues

Reminding and adherence:

- Basic reminders should be provided for blood glucose testing and other relevant activities.
- More advanced features may include reminders when tests are missed, motivational messages to acknowledge achievements, and rewards linked to milestones.
- Development considerations: the aim of developing such features should be to enable dynamic and intelligent tracking of women's behaviors, using reminders and notifications to help prevent unhealthy behaviors, support lifestyle adjustment, and encourage women's behavior change.

Customization:

- Recommended features: demographic data recording, personalized measurement units, customizable normal data ranges, multilanguage options, customizable self-monitoring goals, and a customizable data visualization dashboard.
- Development considerations: the diverse needs, demographic background, and levels of literacy of the women should be considered in app development.

Technical issues:

- Ensure the data details are accurately captured in the app, including numerical precision (for example, decimals), measurement units, and time of data entry.
- Ensure data visualizations are clear and well-organized, and that every recorded data point can be easily viewed.
- Minimize or eliminate connectivity issues and app crashes.

Strengths and Limitations

The strengths of this review include its standardized process and comprehensiveness of evaluation. Standardized evaluation tools were used to measure the quality and multifunctionality of mHealth apps for GDM. In addition, it extends the TECH framework by Gasteiger et al [26] by introducing a more context-based, user-informed, and evidence-informed approach, integrating BCTs, NICE recommendations, and women's preferences, enabling a detailed evaluation of app multifunctionality.

However, several limitations should be considered when interpreting the findings. First, the search was conducted only once on a single day. While this minimizes duplication of included apps, it may have resulted in an incomplete capture of all available apps due to the dynamic nature of app store listings, which can be influenced by advertising and other factors. Second, the decision to restrict the search to UK app stores, while ensuring consistency with UK-based NICE guidelines, may have limited the scope and breadth of the review by excluding insights from apps available in other regions. Third, our findings reflect the status of apps in June 2024. Given the high turnover and frequent updates of apps, their scores, features, and technical issues may have changed postreview, which may affect reproducibility.

Conclusion

This study explored the quality and multifunctionality of commercially available apps for GDM self-management, focusing on the cash-free features, following the TECH and PRISMA frameworks. Our findings highlight that GDM apps should provide more interactive features and support comprehensive pregnancy care, rather than being limited to glucose management and using basic digital techniques. Valuable insights can be drawn from other models of remote pregnancy care, especially those implemented under unavoidable circumstances, such as in rural settings or during the COVID-19 pandemic, where integration of data was required to collect from remote monitoring approaches [47,48]. On one hand, when designing digital technologies for a specific pregnancy complication, it is important to integrate general

pregnancy-related data, such as self-reported symptoms, so that they can be managed alongside the target condition [47]. On the other hand, efforts should be made to enable the in-person services and advice to be available online, ensuring that women receive adequate support while engaging in self-management [48]. Taken together, the evidence highlights an emerging research and design agenda for future mHealth app design—one that prioritizes the integration of comprehensive pregnancy progression management and data interconnectivity to better meet the complex requirements of GDM management.

Acknowledgments

We would like to thank Dr Helen Hawley-Hague for generously lending us the required device.

Authors' Contributions

QZ designed this study, with the support of DD and AC. QZ, YT, and LH reviewed the apps. QZ analyzed the data and wrote the first draft of this paper. QZ, DD, and AC participated in the interpretation of this study's findings and the preparation of this paper. All of the authors approved the final paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Eligibility criteria guided by the TECH framework. TECH: target user, evaluation focus, connectedness, and health domain.
[DOCX File, 19 KB - mhealth_v14i1e76862_app1.docx]

Multimedia Appendix 2
Evidence of evaluation lists.
[XLSX File, 17 KB - mhealth_v14i1e76862_app2.xlsx]

Multimedia Appendix 3
Enlarged PRISMA diagram. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.
[PDF File, 60 KB - mhealth_v14i1e76862_app3.pdf]

Multimedia Appendix 4
Quality appraisal.
[XLSX File, 12 KB - mhealth_v14i1e76862_app4.xlsx]

Multimedia Appendix 5
GDM-adapted features and techniques. GDM: gestational diabetes.
[XLSX File, 20 KB - mhealth_v14i1e76862_app5.xlsx]

Multimedia Appendix 6
General features.
[XLSX File, 14 KB - mhealth_v14i1e76862_app6.xlsx]

References

1. Martis R, Crowther CA, Shepherd E, Alsweiler J, Downie MR, Brown J. Treatments for women with gestational diabetes mellitus: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2018 Aug 14;8(8):CD012327. [doi: 10.1002/14651858.CD012327.pub2] [Medline: 30103263]
2. Tzotzis L, Hooper ME, Douglas A, et al. The needs and experiences of women with gestational diabetes mellitus from minority ethnic backgrounds in high-income nations: a systematic integrative review. Women Birth 2023 Mar;36(2):205-216. [doi: 10.1016/j.wombi.2022.08.006] [Medline: 36038477]
3. Karavasileiadou S, Almegwely W, Alanazi A, Alyami H, Chatzimichailidou S. Self-management and self-efficacy of women with gestational diabetes mellitus: a systematic review. Glob Health Action 2022 Dec 31;15(1):2087298. [doi: 10.1080/16549716.2022.2087298] [Medline: 35867537]
4. Doran F, Davis K. Gestational diabetes mellitus in Tonga: insights from healthcare professionals and women who experienced gestational diabetes mellitus. NZ Med J 2010;123(1326):59-67 [FREE Full text]
5. Evenson KR, Barakat R, Brown WJ, et al. Guidelines for physical activity during pregnancy: comparisons from around the world. Am J Lifestyle Med 2014 Mar;8(2):102-121. [doi: 10.1177/1559827613498204] [Medline: 25346651]

6. Nguyen M, Hossain N, Tangri R, et al. Systematic evaluation of canadian diabetes smartphone applications for people with type 1, type 2 and gestational diabetes. Can J Diabetes 2021 Mar;45(2):174-178. [doi: 10.1016/j.jcjd.2020.07.005] [Medline: 33127288]
7. Eberle C, Stichling S. Effects of telemetric interventions on maternal and fetal or neonatal outcomes in gestational diabetes: systematic meta-review. JMIR Diabetes 2021 Aug 27;6(3):e24284. [doi: 10.2196/24284] [Medline: 34448717]
8. Li SY, Ouyang YQ, Qiao J, Shen Q. Technology-supported lifestyle interventions to improve maternal-fetal outcomes in women with gestational diabetes mellitus: a meta-analysis. Midwifery 2020 Jun;85:102689. [doi: 10.1016/j.midw.2020.102689] [Medline: 32193015]
9. Xie W, Dai P, Qin Y, Wu M, Yang B, Yu X. Effectiveness of telemedicine for pregnant women with gestational diabetes mellitus: an updated meta-analysis of 32 randomized controlled trials with trial sequential analysis. BMC Pregnancy Childbirth 2020 Apr 6;20(1):198. [doi: 10.1186/s12884-020-02892-1] [Medline: 32252676]
10. Global smartphone penetration rate as share of population from 2016 to 2024. Statista. URL: https://www.statista.com/statistics/203734/global-smartphone-penetration-per-capita-since-2005 [accessed 2026-01-20]
11. Khalil C. Understanding the adoption and diffusion of a telemonitoring solution in gestational diabetes mellitus: qualitative study. JMIR Diabetes 2019 Nov 28;4(4):e13661. [doi: 10.2196/13661] [Medline: 31778118]
12. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
13. Zhao Q, Cooke A, Aurizki G, Dowding D. Women's experiences and needs in the use of digital technologies for the management of gestational diabetes: an integrative systematic review. Midwifery 2025 Feb;141:104262. [doi: 10.1016/j.midw.2024.104262] [Medline: 39662131]
14. Safiee L, Rough DJ, Whitford H. Barriers to and facilitators of using eHealth to support gestational diabetes mellitus self-management: systematic literature review of perceptions of health care professionals and women with gestational diabetes mellitus. J Med Internet Res 2022 Oct 27;24(10):e39689. [doi: 10.2196/39689] [Medline: 36301613]
15. Al Hashmi I, Alsabti H, Al Omari O, Al Nasseri Y, Khalaf A. Development, feasibility and acceptability of a self-efficacy-enhancing smartphone application among pregnant women with gestational diabetes mellitus: single-arm pilot clinical trial. BMC Pregnancy Childbirth 2022 Apr 23;22(1):358. [doi: 10.1186/s12884-022-04684-1] [Medline: 35461221]
16. Edwards KJ, Bradwell HL, Jones RB, Andrade J, Shawe JA. How do women with a history of gestational diabetes mellitus use mHealth during and after pregnancy? Qualitative exploration of women's views and experiences. Midwifery 2021 Jul;98:102995. [doi: 10.1016/j.midw.2021.102995] [Medline: 33784541]
17. Safiee L, Rough D, George P, Mudenha R. Baseline perceptions of women with gestational diabetes mellitus and health care professionals about digital gestational diabetes mellitus self-management health care technologies: interview study among patients and health care professionals. JMIR Hum Factors 2023 Dec 19;10:e51691. [doi: 10.2196/51691] [Medline: 38113070]
18. Hirst JE, Mackillop L, Loerup L, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol 2015 Jan;9(1):111-115. [doi: 10.1177/1932296814556506] [Medline: 25361643]
19. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-mum: a feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888. [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
20. Gajecki M, Berman AH, Sinadinovic K, Rosendahl I, Andersson C. Mobile phone brief intervention applications for risky alcohol use among university students: a randomized controlled study. Addict Sci Clin Pract 2014 Jul 2;9(1):11. [doi: 10.1186/1940-0640-9-11] [Medline: 24985342]
21. Akbar S, Coiera E, Magrabi F. Safety concerns with consumer-facing mobile health applications and their consequences: a scoping review. J Am Med Inform Assoc 2020 Feb 1;27(2):330-340. [doi: 10.1093/jamia/ocz175] [Medline: 31599936]
22. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
23. Nouri R, Kalhori SRN, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 1;25(8):1089-1098. [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
24. 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. [doi: 10.2196/mhealth.3422]
25. Kalhori SRN, Hemmat M, Noori T, Heydarian S, Katigari MR. Quality evaluation of english mobile applications for gestational diabetes: app review using Mobile Application Rating Scale (MARS). Curr Diabetes Rev 2021;17(2):161-168. [doi: 10.2174/1573399816666200703181438] [Medline: 32619173]
26. Gasteiger N, Dowding D, Norman G, et al. Conducting a systematic review and evaluation of commercially available mobile applications (apps) on a health-related topic: the TECH approach and a step-by-step methodological guide. BMJ Open 2023 Jun 12;13(6):e073283. [doi: 10.1136/bmjopen-2023-073283] [Medline: 37308269]

27. Cargnin ZA, Schneider DG, Souza MD, Vargas MDO, Tourinho FSV. Low back pain self-management mobile applications: a systematic review on digital platforms. Rev Esc Enferm USP 2024;58:e20230326. [doi: 10.1590/1980-220X-REEUSP-2023-0326en] [Medline: 38875500]
28. Chavez-Ecos FA, Chavez-Ecos R, Vergara Sanchez C, Chavez-Gutarra MA, Agarwala A, Camacho-Caballero K. Mobile health apps for cardiovascular risk assessment: a systematic review. Front Cardiovasc Med 2024;11:1420274. [doi: 10.3389/fcvm.2024.1420274] [Medline: 39376625]
29. Safeer V S M, Gupta P, Behl S, Bansal D, Sahu JK. Mobile health applications for epilepsy in Indian app stores: a systematic review and content analysis using the mobile app rating scale. Epilepsy Res 2024 Mar;201:107331. [doi: 10.1016/j.eplepsyres.2024.107331] [Medline: 38442549]
30. Page MJ, Moher D, Bossuyt PM, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160. [doi: 10.1136/bmj.n160]
31. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR mHealth uHealth 2016 Jun 10;4(2):e72. [doi: 10.2196/mhealth.5849]
32. Aitken M, Gauntlett C. Patient apps for improved healthcare: from novelty to mainstream. : IMS Institute for Healthcare Informatics; 2013 URL: https://ignacioriesgo.es/wp-content/uploads/2014/03/iihi_patient_apps_report_editora_39_2_1.pdf [accessed 2026-01-20]
33. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979;86(2):420-428. [doi: 10.1037/0033-2909.86.2.420]
34. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34. [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
35. Alfawzan N, Christen M, Spitale G, Biller-Andorno N. Privacy, data sharing, and data security policies of women's mHealth apps: scoping review and content analysis. JMIR mHealth uHealth 2022 May 6;10(5):e33735. [doi: 10.2196/33735] [Medline: 35522465]
36. 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. [doi: 10.2196/11130]
37. Diabetes in pregnancy: management from preconception to the postnatal period. NICE. 2015. URL: https://www.nice.org.uk/guidance/ng3 [accessed 2026-01-20]
38. Lef H, Swoboda W, Schobel J, Hieber D, Holl F. Insights into the quality of mobile health apps: preliminary results of an analysis of MARS scores. Stud Health Technol Inform 2024 Aug 22;316:420-421. [doi: 10.3233/SHTI240438] [Medline: 39176767]
39. Tassone C, Keshavjee K, Paglialonga A, Moreira N, Pinto J, Quintana Y. Evaluation of mobile apps for treatment of patients at risk of developing gestational diabetes. Health Informatics J 2020 Sep;26(3):1983-1994. [doi: 10.1177/1460458219896639] [Medline: 31912754]
40. Hayman MJ, Alfrey KL, Waters K, et al. Evaluating evidence-based content, features of exercise instruction, and expert involvement in physical activity apps for pregnant women: systematic search and content analysis. JMIR mHealth uHealth 2022 Jan 19;10(1):e31607. [doi: 10.2196/31607] [Medline: 35044318]
41. Evans K, Donelan J, Rennick-Egglestone S, Cox S, Kuipers Y. Review of mobile apps for women with anxiety in pregnancy: maternity care professionals' guide to locating and assessing anxiety apps. J Med Internet Res 2022 Mar 23;24(3):e31831. [doi: 10.2196/31831]
42. Sahu B, Babu GR, Gurav KS, et al. Health care professionals' perspectives on screening and management of gestational diabetes mellitus in public hospitals of South India - a qualitative study. BMC Health Serv Res 2021 Feb 12;21(1):133. [doi: 10.1186/s12913-021-06077-0] [Medline: 33579259]
43. Utz B, Assarag B, Lekhal T, Van Damme W, De Brouwere V. Implementation of a new program of gestational diabetes screening and management in Morocco: a qualitative exploration of health workers' perceptions. BMC Pregnancy Childbirth 2020 May 24;20(1):315. [doi: 10.1186/s12884-020-02979-9] [Medline: 32448233]
44. Albert L, Capel I, García-Sáez G, Martín-Redondo P, Hernando ME, Rigla M. Managing gestational diabetes mellitus using a smartphone application with artificial intelligence (SineDie) during the COVID-19 pandemic: much more than just telemedicine. Diabetes Res Clin Pract 2020 Nov;169:108396. [doi: 10.1016/j.diabres.2020.108396] [Medline: 32890548]
45. Lu HY, Ding X, Hirst JE, et al. Digital health and machine learning technologies for blood glucose monitoring and management of gestational diabetes. IEEE Rev Biomed Eng 2024;17:98-117. [doi: 10.1109/RBME.2023.3242261] [Medline: 37022834]
46. Yee LM, Leziak K, Jackson J, et al. Patient and provider perspectives on a novel mobile health intervention for low-income pregnant women with gestational or type 2 diabetes mellitus. J Diabetes Sci Technol 2021 Sep;15(5):1121-1133. [doi: 10.1177/1932296820937347] [Medline: 32627582]
47. Kariman SS, van den Heuvel JFM, Adriaanse BME, Oepkes D, Bekker MN. The potential of tele-ultrasound, handheld and self-operated ultrasound in pregnancy care: a systematic review. Prenat Diagn 2025 Jun;45(7):906-920. [doi: 10.1002/pd.6679] [Medline: 39390612]

48. Moise IK, Ivanova N, Wilson C, Wilson S, Halwindi H, Spika VM. Lessons from digital technology-enabled health interventions implemented during the coronavirus pandemic to improve maternal and birth outcomes: a global scoping review. BMC Pregnancy Childbirth 2023 Mar 20;23(1):195. [doi: 10.1186/s12884-023-05454-3] [Medline: 36941565]

Abbreviations

BCT: behavior change technique
GDM: gestational diabetes
ICC: interclass correlation coefficient
IMS: The IMS Institute for Healthcare Informatics Functionality Score
MARS: Mobile App Rating Scale
mHealth : mobile health
NICE: National Institute for Health and Care Excellence
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TECH: target user, evaluation focus, connectedness, health domain

Edited by M DeJonckheere; submitted 09.May.2025; peer-reviewed by C Obianyo, DO Alabi, F Carrouel, M Gasmi, PD Haghighi, S Mohanadas; revised version received 03.Dec.2025; accepted 03.Dec.2025; published 05.Feb.2026.

Please cite as:
Zhao Q, Cooke A, Huang L, Tang Y, Dowding D
Quality and Multifunctionality in Mobile Apps for Gestational Diabetes: Systematic App Review
JMIR Mhealth Uhealth 2026;14:e76862
URL: https://mhealth.jmir.org/2026/1/e76862
doi:10.2196/76862

Original Paper

Evidence of Efficacy of the My Personal Health Guide Mobile Phone App on Antiretroviral Therapy Adherence Among Young African American Men Who Have Sex With Men at 1 Month: Randomized Controlled Trial

Mark S Dworkin[1], MD, MPH&TM; Kara Herrera[1], MPH, PhD; Sierra Upton[1], MS, MPH; Casey M Luc[1], MPH, PhD; Jeb Jones[2], PhD; Paul Burns[3], PhD; Li Liu[1], PhD; Antonio Jimenez[1], PhD; Ruiqi Ren[1], MS; Meaghan Woody[2], MPH; Robert Garofalo[4], MD, MPH; Sangyoon Lee[5], PhD

[1]Community Outreach Intervention Projects, University of Illinois Chicago, Chicago, IL, United States

[2]Rollins School of Public Health, Emory University, Atlanta, United States

[3]John D. Bower School of Population Health, University of Mississippi Medical Center, Jackson, MS, United States

[4]Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, United States

[5]Department of Computer Science, Connecticut College, New London, CT, United States

Corresponding Author:
Kara Herrera, MPH, PhD
Community Outreach Intervention Projects
University of Illinois Chicago
1603 West Taylor Street
Chicago, IL, 60607
United States
Phone: 1 (312) 996 2028
Email: knitti2@uic.edu

Abstract

Background: Young African American men who have sex with men (AAMSM) experience disproportionately high HIV incidence and are less likely to achieve viral suppression compared to White men who have sex with men, an outcome that relies on antiretroviral therapy (ART) adherence. We created My Personal Health Guide, a talking relational agent–based mobile health app to improve ART adherence among young AAMSM.

Objective: The objective was to determine the efficacy of My Personal Health Guide on improving ART adherence among young AAMSM living with HIV.

Methods: We implemented a randomized controlled trial among young (aged 18-34 years) AAMSM with nonoptimal ART adherence throughout the United States between February 2020 and September 2023, predominantly through social media and by word of mouth, provider referral, and fliers in selected health care settings. Participants were randomized in a 1:1 ratio using permuted blocks of 8 to the intervention, My Personal Health Guide, or the attention control arm. ART adherence was assessed with Wilson's 3-item self-reported adherence measurement and dichotomized at ≥80%. Logistic regression models using backward selection were used to evaluate the efficacy of My Personal Health Guide on ≥80% ART adherence at 1-month follow-up.

Results: Among the 253 AAMSM at baseline, most (n=180, 71.1%) self-reported being ≥80% adherent to ART, over half (n=145, 57.3%) resided in the Southern United States, but all US regions were represented, nearly half (n=175, 42.3%) had some college education, over one-third (n=96, 37.9%) had less than optimal literacy, and approximately one-quarter (n=61, 24.1%) experienced housing insecurity in the past 6 months. The sample for analysis of the My Personal Health Guide app efficacy was 131 (intervention=76 and control=55). The odds of being ≥80% adherent to ART at 1-month follow-up were 3.97 (95% CI 1.26-12.55) times greater among participants randomized to the My Personal Health Guide app compared to the controls, after adjusting for ART adherence at baseline, treatment adherence self-efficacy, and ever being incarcerated. Additionally, for every 1-point increase in the HIV Treatment Adherence Self-Efficacy Scale, the odds of ≥80% ART adherence increased by 3% (odds ratio 1.03, 95% CI 1.00-1.06).

Conclusions: Participants randomized to receive My Personal Health Guide reported nearly 4 times greater odds of being ≥80% adherent to ART compared to the attention control group at 1-month follow-up. To our knowledge, this is the first randomized controlled trial demonstrating improved medication adherence using a relational agent–based behavioral intervention. These findings provide evidence of short-term efficacy of My Personal Health Guide to improve ART adherence among young AAMSM. We recommend further research on the inclusion of relational agents in behavioral research, especially in populations affected by stigma and nonoptimal health literacy, where this nonhuman supportive and educational approach may be complementary to health care systems.

Trial Registration: ClinicalTrials.gov NCT04217174; https://clinicaltrials.gov/study/NCT04217174

(JMIR Mhealth Uhealth 2026;14:e75005) doi:10.2196/75005

KEYWORDS

men who have sex with men; African American men who have sex with men; avatar-based mobile phone intervention; Black or African American; HIV; medical adherence; mobile apps; sexual and gender minorities; treatment adherence

Introduction

African American men who have sex with men (AAMSM) have disproportionate HIV infection rates, retention in care [1], mortality, and rates of viral nonsuppression [2-8], an outcome that relies on antiretroviral therapy (ART) adherence. Additionally, health literacy is a factor associated with ART adherence [9] that is lower among African American people than White people [10]. Therefore, interventions that bolster health literacy might improve healthy HIV-related behaviors like ART adherence, potentially resulting in personal and public health benefits.

Relational agents are computerized images that a person may react to as if they are in a relationship. A relational agent can be designed to promote healthy behavioral change. We developed My Personal Health Guide, an innovative theory-driven relational agent–based app with improvement of ART adherence as one of its objectives [11]. During iterative app development focus groups with young AAMSM living with HIV, participants reported universal acceptability for the African American realistic avatar we developed that spoke about, taught, explained, and encouraged behavior to promote healthy HIV-related outcomes [11]. Responding to meeting their adherence-related needs and interests, the app includes clear and simple language, an audiovisual format, and gamification. A pilot study performed in Chicago provided evidence that My Personal Health Guide has the potential to improve adherence [12]. Here, we present preliminary short-term follow-up results from a randomized controlled trial (RCT) of My Personal Health Guide that demonstrate evidence of improvement of ART adherence among young AAMSM living with HIV. We hypothesized that those randomized to the My Personal Health Guide app would have greater ART adherence compared to controls at 1-month follow-up post randomization. Evidence of short-term improvement in ART adherence provides scientific rationale for further research of this realistic avatar and relational agent approach as an adherence intervention and informs efforts to refine the approach to produce longer-term effects.

Methods

Intervention

My Personal Health Guide features a realistic talking avatar with supportive functions that were designed to inform users of HIV-related health information, encourage participants to take their medication, and facilitate adherence-related behavioral skills with reminders and push notifications to take medication and promotion of using weekly pill boxes. It also includes recordings of peers and health care providers offering personal and motivational messages. The app provides a private space, in one's own mobile phone, for hearing and seeing information relevant to healthy living with HIV. Users may replay information, which can be especially helpful to people who are uncomfortable asking questions of their health care provider, including asking for a repeat of what was said or an explanation in simple terms.

Design, Recruitment, Eligibility, and Enrollment

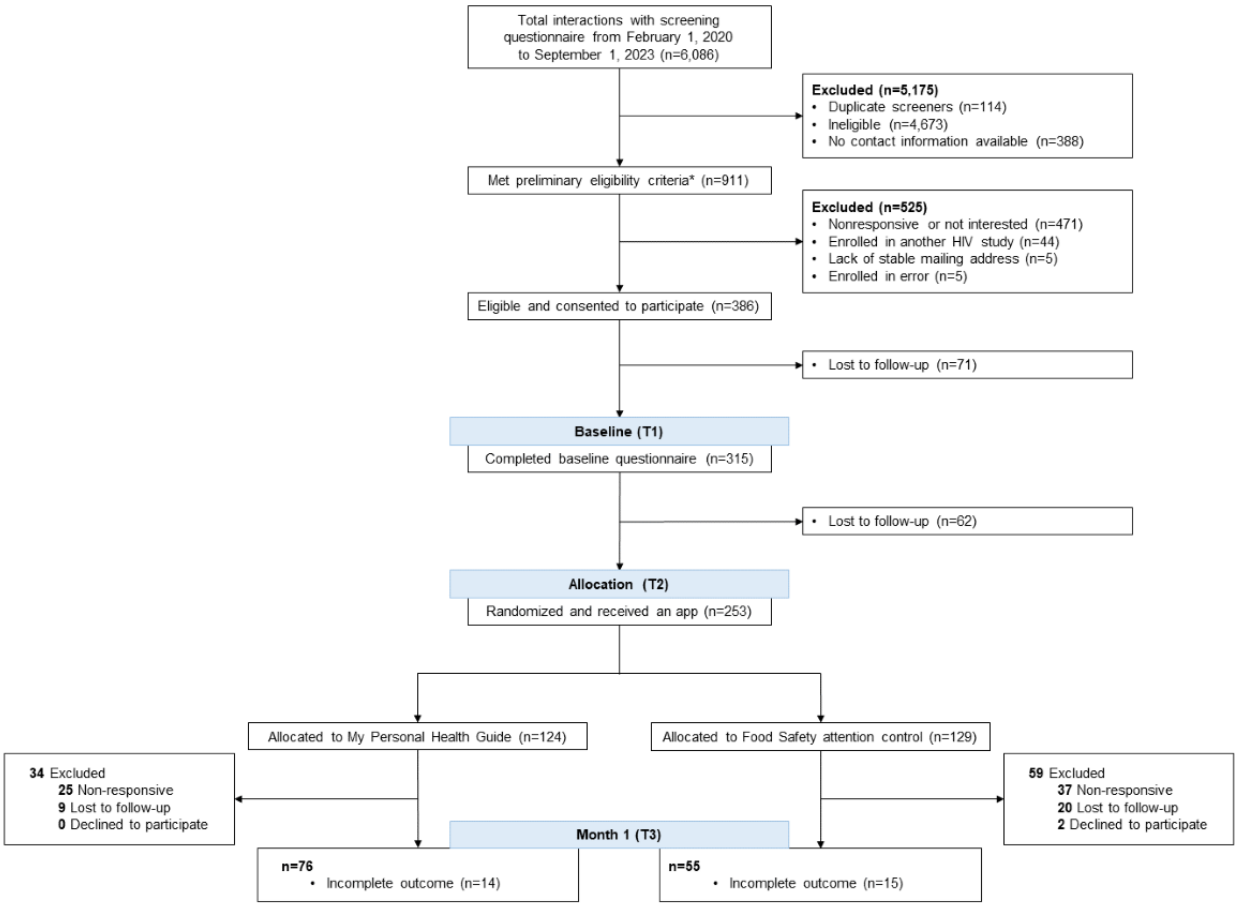
This RCT was performed among young AAMSM living with HIV. Participants were recruited throughout the United States between February 2020 and September 2023, predominantly through social media (eg, the dating app Jack'd and Facebook advertisements) and also by word of mouth, provider referral, and fliers in health care settings of the 3 enrollment sites: University of Illinois at Chicago, Emory University, and University of Mississippi Medical Center. Those who were eligible were assigned male sex at birth, currently identified as a man, African American or Black, aged 18-34 years, had at least one male sexual partner in their lifetime, English speaking, owned a smartphone, were initiating or prescribed oral ART, and had a history of nonoptimal ART adherence defined as having a detectable viral load within the past 4 weeks, self-reported nonoptimal ART adherence within the past 30 days, or were referred by a provider for adherence concerns. Those who were on injectable ART, participating in another HIV treatment study, or did not have a stable mailing address at the time of enrollment were ineligible for participation. This study was determined to be minimal risk, with no adverse events reported.

Data Collection, Sample Size, and Randomization

Figure 1 summarizes participant interactions with the study through the first month of follow-up. A structured,

interviewer-administered baseline questionnaire included collection of demographics, baseline knowledge of HIV (items taught in the app), self-reported adherence, incarceration history, housing stability, substance use history, self-efficacy (HIV Treatment Adherence Self-Efficacy Scale [HIV-ASES]) [13], literacy (adapted Rapid Estimate of Adult Literacy in Medicine-Short Form [REALM-SF]) [14], perceived social support (Multidimensional Scale of Perceived Social Support [MSPSS]) [15], and depression symptomatology (Patient Health Questionnaire for Depression [PHQ-9]) [16]. If a participant exhibited moderately severe or severe depressive symptomatology, they were encouraged to reach out to their provider for mental health resources. One month after the baseline interview, participants were scheduled for the app orientation and download that included explaining how to download, setting up a password, briefly pointing out features, and encouraging app use. Additionally, a check-in call was made monthly during 6 months of follow-up that asked troubleshooting questions about the app and collected self-reported ART adherence. Participants received compensation of US $40 for the baseline interview, US $75 for an appointment to download the app, and US $10 for responding to each monthly check-in call. Sample size calculations were carried out using Repeated Measures and Sample Size (RMASS2), the power calculation software for repeated measures design that allows attrition over time. The study was designed with a proposed sample size of 250 at the end of the study duration, or 295 at baseline, which would provide 80% statistical power to detect a 10% group difference in HIV care retention at the end of the study, considering a 15% attrition rate. Permuted randomization in blocks of 8 was used to allocate research participants in a 1:1 ratio to the intervention or control arm, and allocation concealment from the principal investigator, research staff, and research participants was carried out by the senior statistician. Participants randomized to the intervention arm received the My Personal Health Guide app, and those randomized to the attention control arm received a food safety app, which featured an avatar and functions related to food and nutrition, narrated illustrated stories about food safety, an educational game about sugar content in food, and information promoting oral health. The principal investigator and research participants were blinded after assignment to the intervention. Research coordinators oriented the participants to their respective app after randomization.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram. Preliminary eligibility did not include verifying age and active HIV prescription.



Outcome Measurement

ART adherence was assessed with Wilson's 3-item self-reported adherence measurement, which has been validated against electronic drug monitoring for HIV medication (Cronbach α=0.83) [17]. These three questions include (1) "In the last 30 days, on how many days did you miss at least one dose of any of your HIV medication?" (with a response range from 0 to 30 days); (2) "In the last 30 days, how often did you take your HIV medication in the way you were supposed to?" (with response

options including Never, Rarely, Sometimes, Usually, Almost Always, and Always); and (3) "In the last 30 days, how good a job did you do at taking your HIV medications in the way you were supposed to?" (with response options including Very Poor, Poor, Fair, Good, Very Good, and Excellent). Answers were averaged together and converted to a percentage score on a 100-point scale. Consistent with other studies [18,19], 80% was defined as the cut point for nonoptimal ART adherence. HIV viral load was the intended primary outcome for this research. However, the trial was disturbed by the COVID-19 pandemic, and significant difficulty occurred collecting viral load from this population. Therefore, self-reported ART adherence was used.

Statistical Analysis

The primary statistical goal for this analysis was to determine the efficacy of the My Personal Health Guide app on ART adherence at 1-month follow-up. Participant demographic and study characteristics are presented as means and SDs for continuous variables and as absolute numbers (n) and proportions (%) for categorical variables stratified by randomization arm. The success of randomization was assessed using 2-tailed *t* tests for continuous baseline characteristics or chi-square tests for categorical characteristics. Baseline characteristics between individuals that completed both the baseline meeting and 1-month follow-up versus those who had missing data at 1-month follow-up were compared to ensure comparability between the samples. Bivariate comparisons were made between participant characteristics and ART adherence at baseline (≥80% vs <80%) using chi-square tests for categorical data, Cochran-Armitage tests for ordinal data, and independent *t* tests for continuous data. Logistic regression models with the backward model selection method were used to evaluate the effect of the My Personal Health Guide app, adjusted for significant factors that were associated with ART adherence at month 1, with a $P \leq .05$ considered statistically significant. Baseline adherence was adjusted for a priori. The baseline characteristics that were considered in the logistic regression models included age, ethnicity, remote participation, residence, education, employment status, sexual orientation, currently in a committed relationship, health literacy (REALM-SF), HIV-ASES, depression symptomatology (PHQ-9), perceived social support (MSPSS), ever incarcerated, housing insecurity, and substance use seriousness (serious includes cocaine, methamphetamines, crack, heroin, and opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, and gamma-hydroxybutyrate; none includes no use of substances or alcohol) at baseline. Adjusted model results are presented as odds ratios (ORs) and 95% CI.

Ethical Considerations

This research was reviewed and approved by the institutional review boards of the University of Illinois Chicago, Emory University, and the University of Mississippi Medical Center (protocol #2019-1184). The research involving human data was conducted in accordance with institutional guidelines of the University of Illinois Chicago, Emory University, and the University of Mississippi Medical Center. All participants were administered written informed consent. All data were deidentified prior to analysis and dissemination. Participants were compensated up to US $125 by the 1-month follow-up by their preferred method of check, cash, or e-code.

Results

Sample Characteristics

We attempted to contact the 911 men that met preliminary eligibility criteria, responded to recruitment methods, and provided contact information. Among those, 471 were either not responsive or not interested in the study, 44 were enrolled in another HIV treatment study, and 5 did not have a stable mailing address in order to receive study materials. Five participants were erroneously enrolled in the study due to either being unable to verify having a current HIV prescription (n=3) or having been previously enrolled in the study (n=2). A total of 386 men were consented. Among these, the baseline questionnaire was completed by 315 (81.6%), and 253 (65.5%) had the study apps installed on their mobile phones 1 month later (Figure 1), including 124 (49%) randomized to My Personal Health Guide and 129 (51%) randomized to the Food Safety attention control app. After 1 month, 93 participants did not respond to contact attempts to complete the month 1 check-in call and were excluded from the current analysis. The remaining 160 participants completed the first check-in phone call, and of those, 131 had complete self-reported adherence data at month 1.

The mean age of the 253 participants at baseline was 29.4 (SD 3.7; range18-35) years; most (n=221, 87.4%) were 25-34 years of age. Few (n=18, 7%) identified as Hispanic. Most participated remotely (n=202, 79.8%). Most participants were enrolled by the University of Illinois at Chicago (n=131, 51.8%) and Emory University (n=106, 41.9%), and 16 (6%) were enrolled by the University of Mississippi Medical Center. Over half (n=145, 57.3%) of participants resided in the Southern United States, but all US regions were represented in the sample. Nearly half (n=175, 42.3%) had some college education, and over half (n=146, 57.7%) reported being employed. Three-quarters (n=180, 71.1%) identified as homosexual or gay and were not currently in a committed relationship with a male partner (n=198, 78.3%). Over one-third (n=96, 38%) had less than optimal literacy (adapted REALM-SF). Over half reported mild-severe depressive symptomatology, with 26% (n=65) reporting mild (PHQ-9=5.0-9.9), 14% (n=35) reporting moderate (PHQ-9=10.0-14.9), 9% (n=23) reporting moderately severe (PHQ-9=15.0-19.9), and 4% (n=11) reporting severe (PHQ-9>20.0) depressive symptomatology. The median HIV-ASES score was 104 (IQR 89-114). Although most (n=150, 59.3%) reported high perceived social support (MSPSS>5.0), 9% (n=23) had low (MSPSS<3.0) perceived social support, and 32% (n=80) had moderate (MSPSS=3.0-5.0) perceived social support. Approximately one-quarter experienced housing insecurity in the past 6 months (n=61, 24%), and nearly one-third were ever incarcerated (n=77, 30%). Some (n=39, 15%) reported serious substance use in the past 2 weeks. Nearly three-quarters (n=178, 70.4%) reported less serious substance use, including alcohol, marijuana, ecstasy, speedballs, or gamma-hydroxybutyrate. Most (n=180, 71.1%) self-reported

being ≥80% adherent to ART at baseline. Group comparisons for all baseline characteristics and ART adherence at baseline (Table 1) revealed that randomizations were successful.

Table 1. Baseline demographic characteristics and antiretroviral therapy (ART) adherence of young African American men who have sex with men living with HIV stratified by randomization arm (n=253), 2020-2024.

	Overall (n=253)	Intervention (n=124)	Control (n=129)	*P* value
Age in years, mean (SD)	29.4 (3.7)	29.6 (3.5)	29.1 (4)	.46[a]
Age group (years), n (%)				.41[b]
18-24	32 (12.6)	13 (10.5)	19 (14.7)	
25-34	221 (87.4)	111 (89.5)	110 (85.3)	
Ethnicity, n (%)				>.99[b]
Hispanic	18 (7.1)	9 (7.3)	9 (7)	
Not Hispanic	235 (92.9)	115 (92.7)	120 (93)	
Enrollment site, n (%)				.18[b]
University of Illinois at Chicago	131 (51.8)	59 (47.6)	72 (55.8)	
Emory University	106 (41.9)	54 (43.5)	52 (40.3)	
University of Mississippi Medical Center	16 (6.3)	11 (8.9)	5 (3.9)	
Remote participation, n (%)				.87[b]
Yes	202 (79.8)	98 (79)	104 (80.6)	
No	51 (20.2)	26 (21)	25 (19.4)	
Residence, n (%)				.42[c]
Midwest	77 (30.4)	34 (27.4)	43 (33.3)	
Northeast	27 (10.7)	12 (9.7)	15 (11.6)	
South	145 (57.3)	77 (62.1)	68 (52.7)	
West	4 (1.6)	1 (0.8)	3 (2.3)	
Education, n (%)				.59[b]
College	68 (26.9)	30 (24.2)	38 (29.5)	
Some college	107 (42.3)	51 (41.1)	56 (43.4)	
High school or GED[d]	65 (25.7)	36 (29)	29 (22.5)	
Less than high school	13 (5.1)	7 (5.6)	6 (4.7)	
Employment, n (%)				.45[b]
Active duty	1 (0.4)	0 (0)	1 (0.8)	
Employed	146 (57.7)	75 (60.5)	71 (55)	
Student	18 (7.1)	10 (8.1)	8 (6.2)	
Unable to work or unemployed	83 (32.8)	36 (29)	47 (36.4)	
Missing	5 (2)	3 (2.4)	2 (1.6)	
Sexual orientation, n (%)				.81[b]
Homosexual or gay	180 (71.1)	85 (68.5)	95 (73.6)	
Bisexual	49 (19.4)	27 (21.8)	22 (17.1)	
Heterosexual or straight	4 (1.6)	2 (1.6)	2 (1.6)	
Other	20 (7.9)	10 (8.1)	10 (7.8)	
Currently in a committed relationship with a male partner, n (%)				.66[b]
Yes	55 (21.7)	25 (20.2)	30 (23.3)	
No	198 (78.3)	99 (79.8)	99 (76.7)	

	Overall (n=253)	Intervention (n=124)	Control (n=129)	*P* value
Marital status, n (%)				.69[b]
Legally married	10 (4)	4 (3.2)	6 (4.7)	
Registered domestic partnership	1 (0.4)	1 (0.8)	0 (0)	
Widowed	1 (0.4)	0 (0)	1 (0.8)	
Divorced	8 (3.2)	5 (4)	3 (2.3)	
Separated	5 (2)	2 (1.6)	3 (2.3)	
Never married	228 (90.1)	112 (90.3)	116 (89.9)	
Health literacy (adapted REALM-SF[e]), n (%)				>.99[b]
Nonoptimal (at least 1 wrong)	96 (37.9)	47 (37.9)	49 (38)	
Optimal (none wrong)	157 (62.1)	77 (62.1)	80 (62)	
PHQ-9[f] score, mean (SD)	6.67 (5.81)	6.71 (5.96)	6.62 (5.69)	.90[a]
PHQ-9 score, n (%)				.72[g]
None or minimal (<5.0)	117 (46.2)	55 (44.4)	62 (48.1)	
Mild (5.0-9.9)	65 (25.7)	35 (28.2)	30 (23.3)	
Moderate (10.0-14.9)	35 (13.8)	14 (11.3)	21 (16.3)	
Moderately severe (15.0-19.9)	23 (9.1)	11 (8.9)	12 (9.3)	
Severe (≥20.0)	11 (4.3)	7 (5.6)	4 (3.1)	
Missing	2 (0.8)	2 (1.6)	0 (0)	
HIV-ASES[h], n (%)				.29[a]
Mean (SD)	98.1 (20.2)	99.5 (18.1)	96.8 (22)	
Median (IQR)	104 (89.0-114)	105 (88.3-114)	103 (89.0-115)	
Missing, n (%)	3 (1.2)	2 (1.6)	1 (0.8)	
MSPSS[i] score, mean (SD)	5.08 (1.36)	5.09 (1.45)	5.08 (1.28)	.98[a]
MSPSS score, n (%)				
Low (<3.0)	23 (9.1)	13 (10.5)	10 (7.8)	.68[b]
Moderate (3.0-5.0)	80 (31.6)	37 (29.8)	43 (33.3)	.81[d]
High (≥5.1)	150 (59.3)	74 (59.7)	76 (58.9)	
Ever incarcerated, n (%)				.07[b]
Yes	77 (30.4)	45 (36.3)	32 (24.8)	
No	176 (69.6)	79 (63.7)	97 (75.2)	
Housing insecurity in previous 6 months, n (%)				.64[b]
Yes	61 (24.1)	32 (25.8)	29 (22.5)	
No	192 (75.9)	92 (74.2)	100 (77.5)	
Substance use seriousness in previous 2 weeks[j], n (%)				.32[b]
None	36 (14.2)	17 (13.7)	19 (14.7)	
Less serious	178 (70.4)	92 (74.2)	86 (66.7)	
Serious	39 (15.4)	15 (12.1)	24 (18.6)	
Adherence to ART (baseline), n (%)				.14[b]
Yes (≥80)	180 (71.1)	94 (75.8)	86 (66.7)	
No (<80)	73 (28.9)	30 (24.2)	43 (33.3)	

	Overall (n=253)	Intervention (n=124)	Control (n=129)	*P* value
Adherence to ART (month 1), n (%)				.04[b]
Yes (≥80)	111 (43.9)	69 (55.6)	42 (32.6)	
No (<80)	20 (7.9)	7 (5.6)	13 (10.1)	
Missing data (month 1), n (%)				.005[b]
Yes	122 (48.2)	48 (38.7)	74 (57.4)	
No	131 (51.8)	76 (61.3)	55 (42.6)	

[a]Independent samples *t* test.

[b]Chi-square test for comparison.

[c]Fisher exact test.

[d]GED: General Educational Development.

[e]REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form.

[f]PHQ-9: Patient Health Questionnaire for Depression.

[g]Cochran-Armitage test.

[h]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

[i]MSPSS: Multidimensional Scale of Perceived Social Support.

[j]Serious includes cocaine, methamphetamines, crack, heroin, opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, gamma-hydroxybutyrate.

Primary Outcome

Among the 253 participants who were randomized and downloaded an app, 131 (51.8%) completed the 1-month follow-up for ART adherence, and 122 (48.2%) did not. A significantly greater proportion of those randomized to the Food Safety app were missing 1-month follow-up data compared to the My Personal Health Guide app (74/129, 57.4% vs 48/124, 38.7%; *P*=.005; Table 2). Additionally, remote participation, employment, substance use, and ART adherence at baseline differed by complete 1-month follow-up data and were therefore considered during multivariable model construction. Bivariate analysis (Table 3) revealed the mean HIV-ASES score (mean 103, SD 15.8 vs mean 86.4, SD 28.5; *P*=.02) was significantly higher in those who had ≥80% adherence to ART at 1 month as compared to those who had <80% adherence at 1 month. The mean depression symptomatology scale (PHQ-9) was moderately significantly lower among those who were ≥80% adherent to ART at 1 month as compared to those who were <80% adherent to ART at 1 month (5.70, SD 4.96 vs 9.00, SD 7.43; *P*=.07). There was also a significantly higher prevalence of being in a committed relationship with a male partner among those who had ≥80% adherence to ART at 1 month as compared to those who had <80% adherence at 1 month (25/111, 22.5% vs 0/20, 0%; *P*=.01).

In multivariable analysis (Table 4), the odds of being ≥80% adherent to ART at 1-month follow-up were almost 4 times greater (OR 3.97, 95% CI 1.26-12.55) among participants randomized to the My Personal Health Guide app as compared to the Food Safety app after adjusting for other characteristics that were also associated with ART adherence at 1 month, including HIV-ASES and having a history of ever being incarcerated. Every 1-point increase in the HIV-ASES increased the odds of ≥80% ART adherence by 3% (OR 1.03, 95% CI 1.00-1.06). A history of ever being incarcerated reduced the odds of ≥80% ART adherence (OR 0.31, 95% CI 0.09-1.06). ART adherence at baseline was marginally associated with ≥80% ART at 1-month follow-up (OR 2.97, 95% CI 0.88-9.99).

Table 2. Baseline demographic characteristics and antiretroviral therapy (ART) adherence of young African American men who have sex with men living with HIV stratified by missing outcome (self-reported ART adherence at 1 month) data (n=253), 2020-2024.

	Overall (n=253)	Missing outcome data (month 1; n=122)	Complete outcome data (month 1; n=131)	*P* value
Age in years, mean (SD)	29.4 (3.7)	29.2 (3.6)	29.5 (3.6)	.51[a]
Age group (years), n (%)				.98[b]
18-24	32 (12.6)	16 (13.1)	16 (12.2)	
25-34	221 (87.4)	106 (86.9)	115 (87.8)	
Ethnicity, n (%)				.37[b]
Hispanic	18 (7.1)	11 (9)	7 (5.3)	
Not Hispanic	235 (92.9)	111 (91)	124 (94.7)	
Enrollment site, n (%)				.28[b]
University of Illinois at Chicago	131 (51.8)	66 (54.1)	65 (49.6)	
Emory University	106 (41.9)	46 (37.7)	60 (45.8)	
University of Mississippi Medical Center	16 (6.3)	10 (8.2)	6 (4.6)	
Remote participation, n (%)				<.001[b]
Yes	202 (79.8)	83 (68)	119 (90.8)	
No	51 (20.2)	39 (32)	12 (9.2)	
Residence, n (%)				
Midwest	77 (30.4)	44 (36.1)	33 (25.2)	.16[b]
Northeast	27 (10.7)	15 (12.3)	12 (9.2)	.14[c]
South	145 (57.3)	61 (50)	84 (64.1)	
West	4 (1.6)	2 (1.6)	2 (1.5)	
Education, n (%)				.29[b]
College	68 (26.9)	26 (21.3)	42 (32.1)	
Some college	107 (42.3)	56 (45.9)	51 (38.9)	
High school or GED[d]	65 (25.7)	33 (27)	32 (24.4)	
Less than high school	13 (5.1)	7 (5.7)	6 (4.6)	
Employment, n (%)				.005
Active duty	1 (0.4)	0 (0)	1 (0.8)	
Employed	146 (57.7)	58 (47.5)	88 (67.2)	
Student	18 (7.1)	12 (9.8)	6 (4.6)	
Unable to work or unemployed	83 (32.8)	49 (40.2)	34 (26)	
Missing	5 (2)	3 (2.5)	2 (1.5)	
Sexual orientation, n (%)				.87[b]
Homosexual or gay	180 (71.1)	84 (68.9)	96 (73.3)	
Bisexual	49 (19.4)	25 (20.5)	24 (18.3)	
Heterosexual or straight	4 (1.6)	2 (1.6)	2 (1.5)	
Other	20 (7.9)	11 (9)	9 (6.9)	
Currently in a committed relationship with a male partner, n (%)				.36[b]
Yes	55 (21.7)	30 (24.6)	25 (19.1)	
No	198 (78.3)	92 (75.4)	106 (80.9)	
Marital status, n (%)				

	Overall (n=253)	Missing outcome data (month 1; n=122)	Complete outcome data (month 1; n=131)	*P* value
Legally married	10 (4)	4 (3.3)	6 (4.6)	.71[b]
Registered domestic partnership	1 (0.4)	0 (0)	1 (0.8)	.83[c]
Widowed	1 (0.4)	0 (0)	1 (0.8)	
Divorced	8 (3.2)	5 (4.1)	3 (2.3)	
Separated	5 (2)	2 (1.6)	3 (2.3)	
Never married	228 (90.1)	111 (91)	117 (89.3)	
Health literacy (adapted REALM-SF[e]), n (%)				.41[b]
Nonoptimal (at least 1 wrong)	96 (37.9)	50 (41)	46 (25.1)	
Optimal (None wrong)	157 (62.1)	72 (59)	85 (64.9)	
PHQ-9[f] score, mean (SD)	6.67 (5.81)	7.16 (6.10)	6.21 (5.51)	.20[a]
PHQ-9 score, n (%)				.78[b]
None or minimal (<5.0)	117 (46.2)	55 (45.1)	62 (47.3)	
Mild (5.0-9.9)	65 (25.7)	29 (23.8)	36 (27.5)	
Moderate (10.0-14.9)	35 (13.8)	18 (14.8)	17 (13)	
Moderately severe (15.0-19.9)	23 (9.1)	12 (9.8)	11 (8.4)	
Severe (≥20.0)	11 (4.3)	7 (5.7)	4 (3.1)	
Missing	2 (0.8)	1 (0.8)	1 (0.8)	
HIV-ASES[g] score				.11[a]
Mean (SD)	98.1 (20.2)	96 (21.1)	100 (19.2)	
Median (IQR)	104 (89-114)	105 (90-115)	104 (89-114)	
Missing, n (%)	3 (1.2)	0 (0)	3 (2.3)	
MSPSS[h] score, mean (SD)	5.08 (1.36)	4.96 (1.39)	5.19 (1.33)	.18[a]
MSPSS score, n (%)				
Low (<3.0)	23 (9.1)	11 (9)	12 (9.2)	.12[b]
Moderate (3.0-5.0)	80 (31.6)	46 (37.7)	34 (26)	.17[i]
High (≥5.1)	150 (59.3)	65 (53.3)	85 (64.9)	
Ever incarcerated, n (%)				.14[b]
Yes	77 (30.4)	43 (35.2)	34 (26)	
No	176 (69.6)	79 (64.8)	97 (74)	
Housing insecurity in previous 6 months, n (%)				.57[b]
Yes	61 (24.1)	27 (22.1)	34 (26)	
No	192 (75.9)	95 (77.9)	97 (74)	
Substance use seriousness in previous 2 weeks[j] , n (%)				.02[b]
None	36 (14.2)	25 (20.5)	11 (8.4)	
Less serious	178 (70.4)	77 (63.1)	101 (77.1)	
Serious	39 (15.4)	20 (16.4)	19 (14.5)	
Adherence to ART (baseline), n (%)				.05[b]
Yes (≥80%)	180 (71.1)	79 (64.8)	101 (77.1)	

	Overall (n=253)	Missing outcome data (month 1; n=122)	Complete outcome data (month 1; n=131)	*P* value
No (<80%)	73 (28.9)	43 (35.2)	30 (22.9)	

[a]Independent samples *t* test.

[b]Chi-square test for comparison.

[c]Fisher exact test.

[d]GED: General Educational Development.

[e]REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form.

[f]PHQ-9: Patient Health Questionnaire for Depression.

[g]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

[h]MSPSS: Multidimensional Scale of Perceived Social Support.

[i]Cochran-Armitage test.

[j]Serious includes cocaine, methamphetamines, crack, heroin, opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, gamma-hydroxybutyrate.

Table 3. Bivariate analysis of demographic characteristics of young African American men who have sex with men living with HIV by antiretroviral therapy (ART) adherence at 1 month (n=131), 2020-2024.

	Overall (n=131)	<80% ART adherence (n=20)	≥80 ART adherence (n=111)	*P* value
Randomization, n (%)				.03[a]
My Personal Health Guide (intervention)	76 (58)	7 (35)	69 (62.2)	
Food safety (control)	55 (42)	13 (65)	42 (37.8)	
Age (years), mean (SD)	29.5 (3.62)	29.9 (2.60)	29.4 (3.78)	.53[b]
Age group (years), n (%)				.13[a]
25-34	115 (87.8)	20 (100)	95 (85.6)	
18-24	16 (12.2)	0 (0)	16 (14.4)	
Ethnicity, n (%)				.59[a]
Hispanic	7 (5.3)	0 (0)	7 (6.3)	
Not Hispanic	124 (94.7)	20 (100.0)	104 (93.7)	
Remote participation, n (%)				.39[a]
Yes	119 (90.8)	17 (85.0)	102 (91.9)	
No	12 (9.2)	3 (15.0)	9 (8.1)	
Residence, n (%)				.63[a]
Midwest	33 (25.2)	7 (35.0)	26 (23.4)	
Northeast	12 (9.2)	2 (10.0)	10 (9.0)	
South	84 (64.1)	11 (55.0)	73 (65.8)	
West	2 (1.5)	0 (0.0)	2 (1.8)	
Education, n (%)				.75[a]
College	68 (26.9)	18 (26.5)	50 (73.5)	
Some college	107 (42.3)	30 (28.0)	77 (72.0)	
High school or GED[c]	65 (25.7)	22 (33.8)	43 (66.2)	
Less than high school	13 (5.1)	3 (23.1)	10 (76.9)	
Employment, n (%)				.27[a]
Employed, student, active duty	95 (72.5)	12 (60.0)	83 (74.8)	
Unemployed or unable to work	34 (26.0)	7 (35.0)	27 (24.3)	
Missing	2 (1.5)	1 (5)	1 (0.9)	
Sexual orientation, n (%)				.44[a]
Homosexual or gay	96 (73.3)	18 (90)	78 (70.3)	
Bisexual	24 (18.3)	2 (10)	22 (19.9)	
Heterosexual or straight	2 (1.5)	0 (0)	2 (1.9)	
Other	9 (6.9)	0 (0)	9 (8.1)	
Currently in a committed relationship with a male partner, n (%)				.01[a]
Yes	25 (19.1)	0 (0)	25 (22.5)	
No	106 (80.9)	20 (100)	86 (77.5)	
Health literacy (adapted REALM-SF[d]), n (%)				>.99[a]
Nonoptimal (at least 1 wrong)	46 (35.1)	7 (35)	39 (35.1)	
Optimal (none wrong)	85 (64.9)	13 (65)	72 (64.9)	

	Overall (n=131)	<80% ART adherence (n=20)	≥80 ART adherence (n=111)	*P* value
PHQ-9[e] score, mean (SD)	6.21 (5.51)	9 (7.43)	5.70 (4.96)	.07[b]
PHQ-9 (missing), n (%)	1 (0.8)	0 (0)	1 (0.9)	.07[b]
HIV-ASES[f] score, mean (SD)	100 (19.2)	86.4 (28.5)	103 (15.8)	.02[b]
HIV-ASES (missing), n (%)	3 (2.3)	0 (0)	3 (2.7)	02[b]
MSPSS[g] score, mean (SD)	5.19 (1.33)	4.81 (1.36)	5.26 (1.32)	.18[b]
Ever incarcerated				.41[a]
Yes	34 (26)	7 (35)	27 (24.3)	
No	97 (74)	13 (65)	84 (75.7)	
Housing insecurity in previous 6 months				.41[a]
Yes	34 (26)	7 (35)	27 (24.3)	
No	97 (74)	13 (65)	84 (75.7)	
Substance use seriousness in previous 2 weeks[h]				.12[a]
Serious	19 (14.5)	6 (30)	13 (11.7)	
Less serious	101 (77.1)	13 (65)	88 (79.3)	
None	11 (8.4)	1 (5)	10 (9)	

[a]Fisher exact test.

[b]Independent samples *t* test.

[c]GED: General Educational Development.

[d]REA LM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form.

[e]PHQ-9: Patient Health Questionnaire for Depression.

[f]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

[g]MSPSS: Multidimensional Scale of Perceived Social Support.

[h]Serious includes cocaine, methamphetamines, crack, heroin, opiates; less serious includes alcohol, marijuana, ecstasy, speedballs, gamma-hydroxybutyrate

Table 4. Multivariable logistic regression for the association between My Personal Health Guide and ≥80% antiretroviral therapy (ART) adherence at 1-month follow-up among young African American men who have sex with men living with HIV (n=131), 2020-2024.

	Unadjusted odds ratio	Adjusted odds ratio (95% CI)	*P* value
Randomization			.02
My Personal Health Guide (intervention)	4.55	3.97 (1.26-12.55)	
Food Safety (control)	Reference	Reference	
ART adherence at baseline			.08
≥80%	3.05	2.97 (0.88-9.99)	
<80%	Reference	Reference	
HIV-ASES[a]	1.04	1.03 (1.00-1.06)	.04
Ever incarcerated			.06
Yes	0.60	0.31 (0.09-1.06)	
No	Reference	Reference	

[a]HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale.

Discussion

Principal Findings

Participants randomized to receive My Personal Health Guide demonstrated nearly 4 times greater odds of being ≥80% adherent to ART compared to participants who received the Food Safety control app at 1-month follow-up. This is a substantial magnitude, and although these are short-term follow-up results, these data provide strong evidence that a

relational agent approach to improving medication adherence is promising.

Our study is the first of its kind in the field of HIV medication adherence. We are aware of only 2 other published RCTs of a relational agent for behavior change. Rubin et al [20] performed an RCT of a relational agent to deliver a brief alcohol intervention and referral to treatment to veterans within VA primary care clinics. Their approach involved 2 meetings with the relational agent, who requested a commitment to change. They reported that their relational agent elicited concern about the consequences of drinking and bolstered motivation to change. Although they did not observe accelerated decline in drinking during 3 months with their intervention compared to their control group, they noted that their intervention group was more likely to receive referrals for follow-up and that this approach allowed for more screening and brief intervention without increasing the burden on clinic staff. Another study by Prochaska et al [21] studied a smartphone intervention called Woebot for substance use disorders that used a therapeutic relational agent delivering cognitive behavioral therapy. At 8 weeks, they reported reduced substance use occasions compared to a waitlist group. These studies demonstrate the potential for a relational agent to influence behavior and underscore the scientific rationale for their application to the field of medication adherence. Relational mobile health interventions may be particularly beneficial for younger (18-24 years) AAMSM because they are more likely to still be developing medication-taking routines and may be especially receptive to interactive mobile technologies that can bolster daily ART adherence (Hightow-Weidman et al [22]). Given that these studies and ours are RCTs, they provide evidence for consideration of relational agents in the design of behavioral interventions, especially involving conditions where stigma about diagnosis or treatment could create hesitancy with face-to-face health care provider education or counseling.

In addition to our study demonstrating a large magnitude efficacy, there are several other strengths and considerations. First, our study began shortly before the COVID-19 pandemic spread throughout the United States, disrupting in-person research with lockdowns and other restrictions. This led to a nearly entirely remote rather than facility-based administration and may have made the meeting with research staff for orientation to the app less interactive than it might have been otherwise. The remote approach could have diminished the participant's feeling of accountability to the study or facilitated participation by reducing the barrier of travel to a study activity. Second, unlike medication or immunization, our app intervention involves a passive approach. In other words, patients are encouraged to use it, but it is their choice how much of the app they experience and how often they open it. The app was designed to have new functions and experiences occur with repeated use to encourage return to the app. However, we have no control to ensure regular app exposure. Despite these issues, we observed a significant effect in achieving the adherence goal. This suggests that the effect might increase further and benefit more patients with additional app refinement and involvement of health care providers or case managers in implementation who could refer the app to their patients and check in with them at future appointments. Third, our study sample was diverse, including a large proportion who experienced housing insecurity, screened for moderate or severe depression, and reported a history of incarceration. Additionally, our sample included an overrepresentation of men living in the Southern United States, a high HIV incidence region where many experience structural barriers to HIV care [23] and where a low-cost, scalable intervention like ours could contribute to achieving higher prevalence of viral suppression and help end the HIV epidemic. Finally, our study also demonstrated an independent effect of self-efficacy on being ≥80 adherent. This finding is consistent with other studies, such as those by Dworkin et al [24], Barclay et al [25], and Waldrop-Valverde et al [26]. However, this is the first study to report this specifically in a population of young AAMSM living with HIV.

Limitations

Limitations of this study included that many participants were recruited using a social media dating app and the responsiveness of participants to the 1-month call was moderate, both of which may limit generalizability. Notably, we struggled to recruit younger (18-24 years) AAMSM; therefore, we recommend future studies use a multimodal recruitment approach. Additionally, missingness of self-reported ART adherence was high, and we observed a higher proportion of those randomized to the control with missing data at month 1. Study procedures were disrupted by the COVID-19 pandemic, necessitating the change from a facility-based trial to online. While this pivot allowed for further recruitment and retention, it compromised the collection of viral load data from participants' clinics, which was initially an outcome of interest. This change also impacted the collection of dried blood spots at the 6-month follow-up, which was initially intended to be conducted on-site and changed to be self-administered in the participant's home to accommodate COVID-19 stay-at-home orders. This resulted in a high proportion of missingness. Additionally, the laboratory methodology for the threshold of detection changed during the study period and used a higher threshold (<839 copies/mL) than the clinical definition of undetectable viral load (<200 copies/mL). Similarly, the usage of Wisepill devices to monitor ART adherence was low, so self-report was used as the primary outcome of interest. Using self-reported ART adherence also had limitations. Self-report can overestimate individual medication-taking behavior and therefore may introduce recall and social desirability bias. However, self-reported adherence is often used to assess ART adherence in both routine care and research settings and has been correlated with viral load and other objective markers of adherence [26]. Last, these findings are limited to the first month of observation. Additional follow-up analysis from this trial is planned to determine evidence of long-term effect.

Conclusion

These findings provide evidence that My Personal Health Guide has short-term efficacy in improving ART adherence among young AAMSM living with HIV and contribute to the emerging field of technology-based behavioral interventions promoting health using relational agents. We recommend further research on the inclusion of relational agents in behavioral research,

especially in populations affected by stigma and nonoptimal health literacy, where this nonhuman supportive and educational approach may be complementary to health care systems.

Acknowledgments

We thank the National Institute of Mental Health for funding this research (R01MH116721). We acknowledge Allison Huske, BS; Autumn Smith, MPH; and Elise Dressel, BS, from the University of Illinois at Chicago, School of Public Health; and Paige Jones, BS, and Blake Acebo, BS, from the University of Illinois at Chicago, School of Medicine, for their contributions to data collection.

The National Institute of Mental Health had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The authors were not precluded from accessing data in the study, and they accept responsibility to submit for publication.

Data Availability

A deidentified dataset, a data dictionary defining each field in the set, the study protocol including the statistical analysis plan, and informed consent forms will be made available upon request and with publication acceptance. Per university protocol, all sensitive and high-risk data will be transmitted via an encrypted protocol, or Secure File Transfer Protocol, to individuals outside of the university with investigator support, approval from the data safety and monitoring board, and a signed data access agreement.

Authors' Contributions

MSD and SL contributed to the conceptualization. MSD, RG, and SL contributed to the funding acquisition. MSD, JJ, AJ, RG, and SL contributed to the methodology. MSD, KH, SU, JJ, and PB contributed to project administration. MSD, JJ, PB, RG, and SL contributed to the resources. MSD, JJ, and PB contributed to the supervision. KH, SU, CML, MW, and SL contributed to the data curation. KH, SU, CML, LL, and RR contributed to the formal analysis. KH, SU, and CML contributed to the investigation. KH, SU, CML, and SL contributed to the software. MSD, KH, SU, CML, JJ, LL, and RR contributed to the validation. KH, SU, and RR contributed to the visualization. MSD, KH, SU, CML, LL, and RR contributed to writing the original draft. All authors reviewed and edited the final manuscript.

Conflicts of Interest

The authors report no conflicts of interest. The authors attest that there was no use of generative artificial intelligence technology in the generation of the text, figures, or other informational content of this manuscript.

Multimedia Appendix 1
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 1410 KB - mhealth_v14i1e75005_app1.pdf]

References

1. Singh S, Bradley H, Hu X, Skarbinski J, Hall HI, Lansky A, Centers for Disease ControlPrevention (CDC). Men living with diagnosed HIV who have sex with men: progress along the continuum of HIV care--United States, 2010. MMWR Morb Mortal Wkly Rep 2014;63(38):829-833 [FREE Full text] [Medline: 25254559]
2. HIV among African Americans. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/pdf/risk_HIV_AAA.pdf [accessed 2024-08-22]
3. Estimates on new HIV infections in the United States. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/nchhstp/newsroom/docs/fact-sheet-on-hiv-estimates.pdf [accessed 2024-08-22]
4. HIV among Black/African American gay, bisexual, and other men who have sex with men. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/hiv/risk/racialethnic/bmsm/facts/ [accessed 2024-08-22]
5. Prejean J, Song R, Hernandez A, Ziebell R, Green T, Walker F, HIV Incidence Surveillance Group. Estimated HIV incidence in the United States, 2006-2009. PLoS One 2011;6(8):e17502 [FREE Full text] [doi: 10.1371/journal.pone.0017502] [Medline: 21826193]
6. Back of the line: the state of AIDS among Black gay men in America 2012. Black AIDS Institute. 2012. URL: http://www.blackaids.org/images/reports/back.pdf [accessed 2024-08-22]
7. Diagnoses of HIV infection in the United States and dependent areas. Centers for Disease Control and Prevention. Atlanta, GA: National Center for HIV/AIDS, Viral Hepatitis, STD. and TB Prevention; 2012. URL: http://www.cdc.gov/hiv/library/reports/surveillance/ [accessed 2025-12-06]

8. Beer L, Oster A, Mattson C, Skarbinski J, Medical Monitoring Project. Disparities in HIV transmission risk among HIV-infected black and white men who have sex with men, United States, 2009. AIDS 2014;28(1):105-114 [FREE Full text] [doi: 10.1097/QAD.0000000000000021] [Medline: 23942058]
9. Drainoni M, Rajabiun S, Rumptz M, Welles SL, Relf M, Rebholz C, et al. Health literacy of HIV-positive individuals enrolled in an outreach intervention: results of a cross-site analysis. J Health Commun 2008;13(3):287-302. [doi: 10.1080/10810730801985442] [Medline: 18569359]
10. The health literacy of America's adults: results from the 2003 national assessment of adult literacy. National Center for Education Statistics. URL: https://nces.ed.gov/pubs2006/2006483.pdf [accessed 2018-02-23]
11. Dworkin M, Chakraborty A, Lee S, Monahan C, Hightow-Weidman L, Garofalo R, et al. A realistic talking human embodied agent mobile phone intervention to promote HIV medication adherence and retention in care in young HIV-positive african american men who have sex with men: qualitative study. JMIR Mhealth Uhealth 2018;6(7):e10211 [FREE Full text] [doi: 10.2196/10211] [Medline: 30064971]
12. Dworkin MS, Lee S, Chakraborty A, Monahan C, Hightow-Weidman L, Garofalo R, et al. Acceptability, feasibility, and preliminary efficacy of a theory-based relational embodied conversational agent mobile phone intervention to promote HIV medication adherence in young HIV-Positive African American MSM. AIDS Educ Prev 2019;31(1):17-37. [doi: 10.1521/aeap.2019.31.1.17] [Medline: 30742481]
13. Johnson MO, Neilands TB, Dilworth SE, Morin SF, Remien RH, Chesney MA. The role of self-efficacy in HIV treatment adherence: validation of the HIV treatment adherence self-efficacy scale (HIV-ASES). J Behav Med 2007;30(5):359-370. [doi: 10.1007/s10865-007-9118-3] [Medline: 17588200]
14. Arozullah A, Yarnold P, Bennett C, Soltysik R, Wolf M, Ferreira R, et al. Development and validation of a short-form, rapid estimate of adult literacy in medicine. Med Care 2007;45(11):1026-1033. [doi: 10.1097/MLR.0b013e3180616c1b] [Medline: 18049342]
15. Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support. J Pers Assess 1990;55(3-4):610-617. [doi: 10.1080/00223891.1990.9674095] [Medline: 2280326]
16. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
17. Hightow-Weidman LB, Fowler B, Kibe J, McCoy R, Pike E, Calabria M, et al. HealthMpowerment.org: development of a theory-based HIV/STI website for young black MSM. AIDS Educ Prev 2011;23(1):1-12 [FREE Full text] [doi: 10.1521/aeap.2011.23.1.1] [Medline: 21341956]
18. Ito KE, Kalyanaraman S, Brown JD, Miller WC. Factors affecting avatar use in a STI prevention CD-ROM. J Adolesc Health 2008;42(2 (Suppl 1)):19. [doi: 10.1016/j.jadohealth.2007.11.052]
19. LeGrand S, Muessig KE, McNulty T, Soni K, Knudtson K, Lemann A, et al. Epic Allies: development of a gaming app to improve antiretroviral therapy adherence among young HIV-positive men who have sex with men. JMIR Serious Games 2016;4(1):e6 [FREE Full text] [doi: 10.2196/games.5687] [Medline: 27178752]
20. Rubin A, Livingston NA, Brady J, Hocking E, Bickmore T, Sawdy M, et al. Computerized relational agent to deliver alcohol brief intervention and referral to treatment in primary care: a randomized clinical trial. J Gen Intern Med 2022;37(1):70-77 [FREE Full text] [doi: 10.1007/s11606-021-06945-9] [Medline: 34145518]
21. Prochaska JJ, Vogel EA, Chieng A, Baiocchi M, Maglalang DD, Pajarito S, et al. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. Drug Alcohol Depend 2021;227:108986 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108986] [Medline: 34507061]
22. Hightow-Weidman LB, LeGrand S, Muessig KE, Simmons RA, Soni K, Choi SK, et al. A randomized trial of an online risk reduction intervention for young black MSM. AIDS Behav 2019;23(5):1166-1177 [FREE Full text] [doi: 10.1007/s10461-018-2289-9] [Medline: 30269231]
23. Carter JW, Flores SA. Improving the HIV prevention landscape to reduce disparities for black MSM in the South. AIDS Behav 2019;23(Suppl 3):331-339 [FREE Full text] [doi: 10.1007/s10461-019-02671-w] [Medline: 31541391]
24. Dworkin MS, Chakraborty A, Zychowski D, Donenberg G, Novak R, Garofalo R. Self-efficacy and ability to read as factors associated with antiretroviral therapy adherence in an HIV-infected population. Int J STD AIDS 2018;29(12):1154-1164. [doi: 10.1177/0956462418776073] [Medline: 29890903]
25. Barclay TR, Hinkin CH, Castellon SA, Mason KI, Reinhard MJ, Marion SD, et al. Age-associated predictors of medication adherence in HIV-positive adults: health beliefs, self-efficacy, and neurocognitive status. Health Psychol 2007;26(1):40-49 [FREE Full text] [doi: 10.1037/0278-6133.26.1.40] [Medline: 17209696]
26. Waldrop-Valverde D, Dong C, Ownby R. Medication-taking self-efficacy and medication adherence among HIV-infected cocaine users. J Assoc Nurses AIDS Care 2013;24(3):198-206 [FREE Full text] [doi: 10.1016/j.jana.2012.05.005] [Medline: 23122904]

Abbreviations

AAMSM: African American men who have sex with men

ART: antiretroviral therapy
HIV-ASES: HIV Treatment Adherence Self-Efficacy Scale
MSPSS: Multidimensional Scale of Perceived Social Support
OR: odds ratio
PHQ-9: Patient Health Questionnaire for Depression
RCT: randomized controlled trial
REALM-SF: Rapid Estimate of Adult Literacy in Medicine-Short Form
RMASS2: Repeated Measures and Sample Size

Edited by L Buis; submitted 28.Mar.2025; peer-reviewed by A Adeoye, ME Heidari; comments to author 13.Jun.2025; revised version received 28.Nov.2025; accepted 01.Dec.2025; published 03.Feb.2026.

Please cite as:
Dworkin MS, Herrera K, Upton S, Luc CM, Jones J, Burns P, Liu L, Jimenez A, Ren R, Woody M, Garofalo R, Lee S
Evidence of Efficacy of the My Personal Health Guide Mobile Phone App on Antiretroviral Therapy Adherence Among Young African American Men Who Have Sex With Men at 1 Month: Randomized Controlled Trial
JMIR Mhealth Uhealth 2026;14:e75005
URL: https://mhealth.jmir.org/2026/1/e75005
doi:10.2196/75005
PMID:

Acceptability of Telehealth as the Default Modality for Multiple Sclerosis Care in Switzerland: Cross-Sectional Study

Sintieh Nchinda Ngek Ekongefeyin[1,2,3], MSc, MD; Paola Daniore[4], PhD; Vasileios Nittas[3], PhD; Stefania Iaquinto[3], MSc; Enriqueta Vallejo-Yagüe[3,5], PhD; Christian P Kamm[6,7], Prof Dr Med; Pasquale Calabrese[8], Prof Dr Med; Claudia Baum[9], Dr med; Claudio Gobbi[10,11], Prof Dr Med; Chiara Zecca[10,11], Prof Dr Med; Andrew Chan[7], Prof Dr Med; Milo Puhan[3], Prof Dr; Viktor von Wyl[1,2,3], Prof Dr

[1]Institute for Implementation Science in Health Care, University of Zurich, Universitatstrasse 84, Zurich, Switzerland

[10]Department of Neurology, Neurocenter of Southern Switzerland, Ospedale Regionale di Lugano, EOC, Lugano, Switzerland

[11]Faculty of Biomedical Sciences, Università della Svizzera Italiana, Lugano, Switzerland

[2]Digital Society Initiative, University of Zurich, Zurich, Switzerland

[3]Epidemiology, Biostatistics, and Prevention Institute, University of Zurich, Zurich, Switzerland

[4]Center for Digital Trust, Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland

[5]Institute of Primary Health Care (BIHAM), Faculty of Medicine, University of Bern, Bern, Switzerland

[6]Neurocenter, Luzerner Kantonsspital, Lucerne, Switzerland

[7]Department of Neurology, Inselspital, University Hospital Bern, University of Bern, Bern, Switzerland

[8]Neuropsychology and Behavioral Neurology Unit, Division of Cognitive and Molecular Neuroscience, University of Basel, Basel, Switzerland

[9]Rehaklinik Zihlschlacht AG, Neurological Rehabilitation Center, Part of the VAMED Group, Zihlschlacht, Switzerland

Corresponding Author:
Sintieh Nchinda Ngek Ekongefeyin, MSc, MD
Institute for Implementation Science in Health Care, University of Zurich, Universitatstrasse 84, Zurich, Switzerland

Abstract

Background: Telehealth can improve access to care for people living with multiple sclerosis (MS), but information on its acceptance is limited in Switzerland.

Objective: This study aimed to determine the proportion of people living with MS willing to accept telehealth as a new default and the factors associated with their acceptance.

Methods: We conducted a cross-sectional analysis using survey data from the Swiss Multiple Sclerosis Registry. We defined "telehealth as a default" as a health care model where remote consultations (telephone and/or video calls) are the primary mode of interaction between patients and their physicians, with in-person visits based on clinical necessity. Multivariable logistic regression was performed to evaluate the association between telehealth acceptance and sociodemographic and health-related factors. Telehealth acceptance was described in relation to 3 survey variables that mirrored key constructs from the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework. The variables were digital communication preferences, internet use for health provider searches, and experience with telemedicine.

Results: Among 427 respondents, 15.5% (66/427) reported a willingness to accept telehealth as their default. In this group, only 21.2% (14/66) had experience using telemedicine. A descriptive analysis of our 3 NASSS-derived key constructs showed that among the 78.5% (335/427) respondents who generally agreed to digital access to health data, only 17.0% (57/335) accepted telehealth as a default. Notably, 30.7% (129/427) of participants stated a wish for support for using devices or the internet. Among those 129 individuals, 17.1% (22/129) were willing to accept telehealth as a default. Of the 89 people with prior telehealth experience, 15.7% (14/89) were willing to accept telehealth. In multivariable analysis, digital communication with health care providers (adjusted odds ratio [aOR] 14.56, 95% CI 6.18 - 39.04; *P*<.001*)*, current internet use for health care provider search (aOR 7.78, 95% CI 1.34 - 45.32; *P*=.021), and a secondary progressive MS diagnosis (aOR 0.22, 95% CI 0.05 - 0.72; *P*=.021) were independently associated with accepting telehealth as a default.

Conclusions: Our findings suggest a low acceptance of telehealth as a default among people living with MS in Switzerland. While our 3 postulated NASSS-derived key constructs were not associated with telehealth acceptance, we noted additional behavioral factors, including previous digital communication with health care providers and using the internet to search for health care provider information, which were associated with telehealth acceptance. Moreover, advanced disease states like secondary progressive MS were negatively associated with telehealth acceptance. Thus, telehealth as a default will be most acceptable in

people living with MS who already use the internet for their health, and those with less severe disease. Future research should explore provider perspectives and evaluate long-term strategies for the acceptance of telehealth in MS care.

(JMIR Mhealth Uhealth 2026;14:e84447) doi:10.2196/84447

KEYWORDS

acceptance; chronic disease management; digital health; multiple sclerosis; patient preferences; telehealth as a default; telemedicine

Introduction

Telehealth, the use of information and communication technology to provide remote health care, is a relatively novel approach in chronic disease management [1]. If widely adopted, it has the potential to improve access, continuity, and quality of care while reducing costs for populations with chronic diseases [2,3]. Evidence of patient benefits through its application in chronic disease care is promising. When used among people with cardiovascular disease (CVD), half of them had improved CVD risk scores [4]. When used among people living with diabetes, the quality of care was comparable to standard care while improving access and reducing the cost of care [5]. Among people living with cancer, it improved the quality of life and psychological well-being [6]. The COVID-19 pandemic also accelerated the use of telehealth in patient care, demonstrating its feasibility and acceptability for both patients and clinicians in chronic disease care, including multiple sclerosis (MS) [7,8].

In MS care, telehealth has been used for follow-up consultations, cognitive assessments, rehabilitation, and the management of symptoms that may limit mobility or access to specialist care. Altman et al [9] reported that both people living with MS and health care providers had a high level of satisfaction with telehealth. A review reported that people living with MS and their health care providers generally viewed telehealth as an acceptable and effective alternative to traditional in-person visits [10]. Also, telehealth has been reported to improve self-management [11,12], anxiety and depression [13-15], and activities of daily living in people living with MS [16]. Finally, no adverse events were reported with telerehabilitation for people living with MS [17].

The successful implementation of telehealth in MS and other chronic diseases is influenced by a range of enablers and barriers. Enablers include the perceived value proposition of the technology, such as improved access, convenience, and patient outcomes, alongside supportive organizational structures and positive attitudes among both patients and clinicians [18-20]. Conversely, barriers often arise from the severity of the disease, organizational complexity, technological challenges, digital literacy gaps, and concerns regarding data security or the adequacy of remote clinical assessments [16,21-23].

To systematically assess enabling and inhibiting factors, several frameworks have been proposed [24-26]. A widely used framework is the Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) framework [27]. The NASSS framework provides a comprehensive approach to understanding the adoption and sustained use of technologies in health care and has been used for a wide range of technologies and disease areas [27], including the use of telehealth in chronic care for conditions like CVDs, cancers, and neurological disorders [26]. The framework evaluates the complexity of technology acceptance within 7 constructs included in NASSS, which aim to identify and address factors influencing technology acceptance, anticipate barriers to implementation, and design interventions that are more likely to achieve scale-up and sustainability.

In MS care, few studies examine under what conditions people living with MS would accept telehealth as a default. In our study, we define “telehealth as a default” as a health care delivery model where remote consultations are the primary mode of interaction between patients and their physicians, with in-person visits occurring based on clinical necessity [28]. Previous studies report high patient and provider satisfaction [29], variable confidence in using telehealth in MS care [19], barriers limiting telehealth adoption [23]; but fail to report what conditions would enable or inhibit telehealth as a default among people living with MS.

Overall, many preconditions for successful telehealth implementation in Switzerland should be in place, including the presumed (and partially demonstrated) advantages of telehealth, the generally positive health-seeking attitude of people living with MS in Switzerland [30,31], and the push from global and national authorities for digital infrastructure to reduce diagnostic and therapeutic delays [1,32,33]. However, it remains unclear to what extent the Swiss population of people living with MS is capable and willing to engage in telehealth as a default mode for routine consultations. Therefore, our research seeks to assess the proportion of people living with MS in Switzerland potentially willing to accept telehealth as a default and the factors influencing their acceptance. Our findings are contextualized using the NASSS framework [34] and will provide information on how to target people living with MS with telehealth and consequently improve the quality of care for people living with MS.

Methods

Study Design and Setting

This study was a nested cross-sectional analysis within the Swiss Multiple Sclerosis Registry (SMSR), an ongoing observational cohort initiated in 2016. The SMSR recruits adults with MS residing or receiving treatment in Switzerland. Participation is voluntary and requires written informed consent, alongside confirmation of MS or clinically isolated syndrome (CIS) diagnosis. Upon enrollment, participants complete a baseline questionnaire and are invited to provide self-reported data via follow-up questionnaires approximately twice a year, available both online and in paper format.

Ethical Considerations

Ethical approval for the SMSR was granted by the ethics committee of the Canton of Zurich (PB–2016–00894; BASEC-NR 2019–01027). Further details on the SMSR have been previously published [35,36]. All participants had previously provided informed consent to be contacted and the use of their data for research purposes, including secondary analyses. Participation in this survey was voluntary, and individuals could decline without consequences. All data provided to the study team were fully anonymized. To protect participants' privacy and confidentiality, only anonymized data were used for this analysis, and results are reported in aggregate such that no individual participant can be identified. Participants received no compensation for completing the survey.

For this analysis, we used data from an anonymous, complementary online survey conducted between April 2023 and June 2023. This survey is a follow-up investigation of an earlier study on patterns of digital device use and digital literacy that was launched in October 2020 [37]. The 2023 survey used for this analysis was conducted online, in German, and offered independently of routine SMSR follow-ups (n=1720 eligible; 431 respondents; Multimedia Appendix 1). The rationale for offering the follow-up survey to a limited subsample had several reasons. First, we aimed to inform ongoing debates on telehealth and electronic health records promptly and therefore decided to offer the survey only in one language and online, also to facilitate data analysis. Second, our decision to offer the survey online and anonymously was also designed to attract inactive participants (ie, those who had not completed surveys in the past 2 years but remained in the cohort). Including these participants was intended to mitigate nonresponse bias and to capture perspectives from people living with MS who may be less engaged with research or digital tools. Third, the well-described SMSR source population and the existence of results from the 2020 study, which was offered to the full SMSR study population in all languages and participation modes (online and paper-pencil), enabled us to examine the generalizability and projections of key findings based on the limited, second survey (Multimedia Appendix 1).

Using the NASSS Framework for Hypothesis Development and Interpretation of Findings

The NASSS framework provides a structured approach to understanding the facilitators and barriers to technology uptake [34]. In our study, we used the NASSS framework to generate hypotheses, given the available data, for which factors may determine the acceptance of telehealth as a default by people living with MS in Switzerland. It comprises 7 domains: *the condition*—the nature of the health condition; *the technology*—the technology itself (eg, usability and reliability); *the value proposition*—the value of the technology to health care stakeholders; *the adopters*—the individuals expected to adopt the technology (patients, clinicians, and carers); *the organization*—the health care organizational context; *the wider system*—the wider system including policy and regulatory factors; and *embedding over time*—how the technology and its use may evolve through processes of embedding and adaptation. To describe the complexity of technology acceptance in the aforementioned domains, the NASSS framework defines a "simple technology" as one that has few components and its successful implementation is predictable, a "complicated technology" as one that has multiple components and its success is largely predictable, and a "complex technology" as one that has numerous interacting components with limited chances of succeeding and hence, less likely to be adopted [27]. Systematic reviews applying the NASSS framework have identified that barriers are most frequently encountered within the domains of organization and adopter system, while enablers are linked to clear value propositions and alignment with patient needs [26,34,38].

Main Outcomes

The primary study outcome was the proportion of participants who accepted telehealth as a default method for consultations. The acceptance of telehealth as a default was assessed using the following question: "Agreement to Digital Default (Would you find it beneficial for your medical care if medical consultations were through telemedicine?)," collected as a Likert scale from 0 (completely disagree) to 4 (completely agree). For further analysis, a binary variable was created where "Yes" included Likert levels (3,4) and "No" included levels (0,1,2). When describing the prevalence of acceptance in the population, we refer to the observed proportion as "low acceptance" descriptively; we did not define formal quantitative thresholds for low versus high acceptance.

Main Explanatory Variables of Interest

We hypothesized 3 main drivers of telehealth acceptance that are included in our data: (1) knowledge of and prior experience with telehealth, (2) general openness toward digital health tools and electronic health records, and (3) needs and wishes for digital support.

Knowledge of and experience with telehealth was assessed using the following question: "Do you have knowledge of telemedicine offerings, and have you already gathered experiences with telemedicine?" Answer options were "neither knowledge of nor experience with telemedicine," "knowledge of but no experience with telemedicine," or "knowledge of and experience with telemedicine."

Openness for use of other digital health tools was assessed through items on willingness to accept other digital health services (eg, file sharing, health data use, and provider communication) and experience with telemedicine. Each variable was collected as a Likert scale ranging from 0 (completely disagree with digital default) to 4 (completely agree with specific digital default).

Digital support needs were assessed with two items: (1) "Do you require support in using digital devices (eg, smartphones, laptops)?" and (2) "Do you need assistance with internet-related tasks (eg, security, navigation, program management)?" Each was rated on a Likert scale from 0 (no support needed) to 2 (extensive support needed). Participants reporting moderate (1) or extensive support (2) needs for both items were classified as requiring digital support.

Other Explanatory Variables

Covariates included age (categorized as 18 - 30, 31 - 40, 41 - 50, 51 - 60, 61 - 70, >70 y), sex (male or female), MS type (CIS or relapsing-remitting MS, primary progressive MS [PPMS], secondary progressive MS [SPMS]), and frequency of internet use for health-related purposes (eg, appointment scheduling and provider communication). Prior or current telemedicine use was included as a binary variable. Additional covariates included concerns related to internet use (eg, data security and information reliability). Internet use frequency was categorized as never, less than monthly, monthly, weekly, daily, or several times daily. A composite "use score" (range 0 - 5) was calculated by summing the number of online activities performed at least monthly.

Statistical Analysis

We first described the study population using MS type, sociodemographic characteristics, and experience with telemedicine. We then compared key variables based on previous telehealth experience (Multimedia Appendix 2) and openness to default telehealth consultations (see Table 1). In this study, a complete case analysis was undertaken. We excluded persons where key variables were missing from the analysis, leading to the exclusion of 4 participants.

Table . Comparison of preferences by default telehealth consultation status (N=427).

Variables	Overall, n (%)	Telehealth consultation as default =no (n=361), n (%)	Telehealth consultation as default=yes (n=66), n (%)
Descriptive variables			
Age group (y)			
18 - 30	11 (2.6)	9 (2.5)	2 (3.0)
31 - 40	56 (13.1)	48 (13.3)	8 (12.1)
41 - 50	108 (25.3)	90 (24.9)	18 (27.3)
51 - 60	143 (33.5)	121 (33.5)	22 (33.3)
61 - 70	81 (19.0)	73 (20.2)	8 (12.1)
>70	28 (6.6)	20 (5.5)	8 (12.1)
Sex			
Female	298 (69.8)	256 (70.9)	42 (63.6)
Male	129 (30.2)	105 (29.1)	24 (36.4)
MS[a] types			
CIS[b] and RRMS[c]	285 (66.7)	240 (66.5)	45 (68.2)
PPMS[d]	51 (11.9)	44 (12.2)	7 (10.6)
SPMS[e]	91 (21.3)	77 (21.3)	14 (21.2)
Hypothesis 1: Have you ever used a telemedicine service?			
No knowledge, no experience	169 (39.6)	143 (39.6)	26 (39.4)
Knowledge, no experience	169 (39.6)	143 (39.6)	26(39.4)
Knowledge, experience	89 (20.9)	75 (20.8)	14 (21.2)
Hypothesis 2: acceptability to digital default (Would you find any of the following functionalities beneficial for your medical care?)			
Sell your health data (agree)	102 (24.2)	72 (20.2)	30 (45.5)
Default digital access to health data (agree)	335 (78.5)	277 (76.7)	58 (87.9)
Default digital document exchange (agree)	236 (55.7)	187 (52.2)	49 (74.2)
Default digital communication with a health care provider (agree)	167 (39.1)	114 (31.6)	53 (80.3)
Hypothesis 3: In which areas do you wish for more support in the “digital” world?			
Need support when using the internet (yes)	104 (24.8)	85 (24.0)	19 (28.8)
Need support when using hardware (yes)	109 (25.8)	86 (24.2)	23 (34.8)
Needs support for both	129 (30.2)	105 (29.7)	24 (36.4)
Other potentially explanatory variables			
Which concerns do you have related to internet use?			
Data security (agree)	271 (64.2)	235 (66.0)	36 (54.5)
Reliability of information “invalid data” (agree)	252 (60.0)	220 (62.1)	32 (48.5)
Isolation from real life (agree)	97 (23.1)	80 (22.6)	17 (25.8)

Variables	Overall, n (%)	Telehealth consultation as default =no (n=361), n (%)	Telehealth consultation as default=yes (n=66), n (%)
Health effects of internet use (agree)	70 (16.6)	54 (15.2)	16 (24.2)
For which activity do you go online when related to the topic of medicine and health (components for use score)? Percentage with at least 1 monthly activity			
Exchange with other patients	41 (10.5)	33 (9.9)	8 (13.6)
Self-tracking	43 (11.2)	32 (9.7)	11 (20.0)
Communication with a health care provider	57 (14.0)	44 (12.8)	13 (21.0)
Health care provider search	30 (7.5)	21 (6.2)	9 (15.0)
MS information search	195 (48.6)	157 (46.3)	38 (61.3)
Health information search	249 (62.7)	209 (61.5)	40 (70.2)
Private appointments (yes)	250 (58.5)	208 (57.6)	42 (63.6)
Professional appointments (yes)	188 (44.0)	157 (43.5)	31 (47.0)
All appointment types (yes)	299 (70.0)	250 (69.3)	49 (74.2)
How often do you use electronic devices (at least monthly)			
Frequency of smartwatch use	57 (18.6)	48 (18.0)	9 (22.0)
Frequency of smartphone use	365 (95.3)	307 (94.8)	58 (98.3)
Frequency of tablet use	146 (45.8)	126 (45.7)	20 (46.5)
Frequency of PC use	294 (89.1)	246 (88.8)	48 (90.6)

[a]MS: multiple sclerosis.

[b]CIS: clinically isolated syndrome.

[c]RRMS: relapsing-remitting multiple sclerosis.

[d]PPMS: primary progressive multiple sclerosis.

[e]SPMS: secondary progressive multiple sclerosis.

Univariable logistic regression models were used to identify factors associated with willingness to accept default telehealth consultations (Multimedia Appendix 3). All prespecified variables relating to the use of digital tools and attitudes toward digitalization were included in the full multivariable logistic regression model, which also included predefined adjustments for potential confounders (age, sex, MS type, and telemedicine experience). A 2-sided *P* value <.05 was considered statistically significant. All analyses were conducted using R (version 4.4.2; R Foundation for Statistical Computing).

Reporting Framework

This paper was written following the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines (Checklist 1) for cross-sectional studies [39].

Results

Sociodemographic Data

Of 431 survey respondents, 4 had missing information regarding age. Therefore, a total of 427 participants with complete data were included in the analysis. The study sample consisted of 30.2% (129/427) men and 69.8% (298/427) women, with 41% (175/427) of participants aged between 18 and 50 years. Of the respondents, 66.7% (285/427) reported having relapsing-remitting MS or CIS, while 11.9% (51/427) had PPMS and 21.3% (91/427) SPMS. A total of 20.9% (89/427) reported previous experience with telemedicine consultations (Table 2).

Table . Characteristics of participants (N=427).

Variable	All participants, n (%)
Age group (y)	
18 - 30	11 (2.6)
31 - 40	56 (13.1)
41 - 50	108 (25.3)
51 - 60	143 (33.5)
61 - 70	81 (19)
70+	28 (6.6)
Sex	
Female	298 (69.8)
Male	129 (30.2)
MS[a] types	
CIS[b] and RRMS[c]	285 (66.7)
PPMS[d]	51 (11.9)
SPMS[e]	91 (21.3)
Experience or knowledge with telemedicine	
No experience or knowledge	169 (39.6)
Telemedicine knowledge, no experience	169 (39.6)
Telemedicine knowledge and experience	89 (20.9)

[a]MS: multiple sclerosis.

[b]CIS: clinically isolated syndrome.

[c]RRMS: relapsing-remitting multiple sclerosis.

[d]PPMS: primary progressive multiple sclerosis.

[e]SPMS: secondary progressive multiple sclerosis.

Acceptance of Telehealth as a Default Method of Consultation

A total of 66 out of 427 (15.5%) participants were open to a default telehealth consultation (Table 1). Among this group, 21.2% (14/66) had previous telemedicine experience. Although 78.5% (335/427) agreed to default digital access to health data, only 17.3 (58/335) agreed to default to telehealth. Of the 30.2% (129/427) participants needing hardware and software support, 81.4% (105/129) did not agree to telehealth as a default. Older age groups greater than or equal to 60 years 90.1% (73/81,) and those concerned with data security 86.7% (235/271,) and information validity 87.3%(220/252,) rarely accepted telehealth as a default.

Factors Associated With Acceptance of Telehealth as a Default

Multivariable Regression

In the multivariable regression model (Table 3), default communication with health care providers (adjusted odds ratio [aOR] 14.56, 95% CI 6.18 - 39.04; *P*<.001), internet use for health care provider search (aOR 7.78, 95% CI 1.34 - 45.32; *P*=.021) and a SPMS diagnosis (aOR 0.22, 95% CI 0.05 - 0.72; *P*=.021) were independently associated with an agreement to default telehealth consultation. Other factors, such as concerns about invalid information, data security, and telemedicine experience, did not show significant associations when adjusted for other predictors. The multivariable model showed low multicollinearity (variance inflation factor <1.5), suggesting that the predictors are not highly correlated with each other and that the regression coefficients can be reliably interpreted.

Table . Multivariable logistic regression for telehealth consultations.

Variable	aOR[a] (95% CI)	*P* value
Descriptive variables		
Sex (male)	0.88 (0.35-2.11)	.778
Age (y)		.271[b]
18 - 30 (reference)	—[c]	
31 - 40	0.49 (0.07-4.60)	
41 - 50	0.60 (0.10-5.12)	
51 - 60	0.67 (0.10-5.89)	
61 - 70	0.55 (0.07-5.41)	
>70	4.15 (0.40-54.92)	
MS[e] type		.041[d]
MS type: CIS[f] or RRMS[g] (reference)	—	
MS type: PPMS[h]	0.67 (0.17-2.24)	
MS type: SPMS[i]	0.22 (0.05-0.72)	
Hypothesis 1: Have you ever used a telemedicine service?		
Telemedicine experience (yes or no; reference)	0.94 (0.37-2.23)	.888
Hypothesis 2: agreement to digital default (Would you find any of the following functionalities beneficial for your medical care?)		
Default digital communication with HCP[j]	14.56 (6.18-39.04)	<.001
Default digital access to health data	0.65 (0.19-2.34)	.492
Default digital document exchange	1.60 (0.66-4.15)	.311
Hypothesis 3: In which areas do you wish for more support in the “digital” world?		
Needs support (yes or no)	1.98 (0.84-4.60)	.112
Other explanatory variables		
Concern about invalid information	0.96 (0.42-2.20)	.920
Concerned about data security	0.85 (0.37-1.97)	.702
Uses the internet for health info search	0.86 (0.37-2.05)	.737
Uses internet for MS info search	1.68 (0.73-3.93)	.225
Uses the internet for HCP search	7.78 (1.34-45.32)	.021

Variable	aOR[a] (95% CI)	*P* value
Uses the internet for communication with an HCP	0.62 (0.19-1.82)	.409
Uses the internet for self-tracking	1.38 (0.45-3.95)	.560
Uses the internet for patient exchange	1.02 (0.28-3.21)	.980

[a]aOR: adjusted odds ratio.

[b]Likelihood ratio test: χ^2_5=5.49; *P*=.36.

[c]Not applicable.

[d]Likelihood ratio test: χ^2_2=6.44; *P*=.040.

[e]MS: multiple sclerosis.

[f]CIS: clinically isolated syndrome.

[g]RRMS: relapsing-remitting multiple sclerosis.

[h]PPMS: primary progressive multiple sclerosis.

[i]SPMS: secondary progressive multiple sclerosis.

[j]HCP: health care professional.

Discussion

Summary of Findings

This cross-sectional study determined the proportion of people living with MS who are willing to accept telehealth as a default and the factors influencing their acceptance. We found that only 66 out of 427 (15.5%) participants were willing to accept telehealth as a default, which was positively associated with a general openness for digital communication with health care providers or using the internet for health care provider searches. We also found a negative association of telehealth acceptance with having SPMS.

The Proportion of People Living With MS Who Accepted Telehealth as a Default

In our study, less than one-fifth of people living with MS accepted telehealth as a default. However, this finding contrasts with other studies reporting over 70% acceptance of telehealth among people living with MS compared to in-person visits [23,29,40]. These earlier studies examined acceptance and satisfaction with telehealth during the COVID-19 pandemic, a period in which the extensive use of telemedicine ensured continuity of care, which probably contributed to increased acceptance. Hence, the difference in rates of acceptance might be explained by the forced global shift to digital services to curb the pandemic. Also, these earlier investigations did not specifically assess accepting telehealth as a default. Our low acceptance rate could also be due to the complexity of MS and factors like comorbidities and the older age of our study sample [41]. MS severity is often linked with more physical impairments, comorbidities, and aging [42]. In our study, persons with PPMS and SPMS, as well as persons aged over 60 years, were overrepresented compared with the projected population of people living with MS in Switzerland (Multimedia Appendix 1). These sample differences may explain the low percentage of telehealth acceptance as a default. Indeed, studies have repeatedly shown that people aged above 60 years are least accepting telehealth as a default; that corroborates with research on technology acceptance in older populations [23,27,43-45]. Nevertheless, some research highlights that older adults may still be receptive to telehealth when it addresses access barriers, reduces the burden of travel, or is perceived as enhancing continuity of care [46,47]; hence, addressing these preconditions may increase acceptance of telehealth.

The contrast between the high acceptance of default digital access to health data (78.5%) and the low acceptance of telehealth as default MS care (17.1%) suggests that people living with MS may favor asynchronous over synchronous digital modalities. Asynchronous tools such as secure messaging, patient portals, and digital care paths enable patients to access information, contact their care team, and receive tailored feedback without the need for real-time interaction and have been shown to support chronic disease management and self-management, including in MS [48-50]. Our findings most likely suggest the use of asynchronous telehealth as a complementary, but not necessarily default, component of chronic MS care.

Factors Associated With Accepting Telehealth as a Default

In our multivariable regression analysis, neither of the 3 prepostulated NASSS constructs as possible drivers of telehealth acceptance (hypotheses 1 - 3) showed independent associations. For example, although prior telemedicine experience has been linked to greater acceptance in other settings, it was not significantly associated with acceptance in our cohort. Only 20% of participants reported any prior telemedicine experience—consistent with the generally low acceptance of telehealth among people living with MS in Switzerland [32]—which likely limited its influence in our analysis.

However, we found that a general acceptance toward digital communication with health care providers, as well as regular internet use to search for health care providers, was positively associated with a willingness to accept telehealth as a default.

A general acceptance of digital communication may indeed ease a transition toward telehealth, accelerated by the COVID-19 pandemic, when tools for digital communication through voice and video became more widely adopted, also among older populations [8,10,29,51]. The positive association of prior experience with online searches of health care providers is more challenging to interpret but may signal a general interest in using the internet as a tool for personal health management [7,11]. In the first digital survey, frequent online searches for health care providers distinguished a subgroup of people living with MS with greater health awareness who regularly used digital tools for managing their condition [37]. In general, individuals familiar with digital tools may perceive a lower cognitive and practical barrier in using telehealth [52]. This observation could also be true for those who use the web to search for care providers. Most likely explaining the positive association between telehealth and prior use of the internet for health care provider search found in our study, because the user's perceived effort with technology will be relatively low [45].

Individuals with SPMS were less likely to endorse telehealth as a default. This may reflect the greater physical and cognitive burden associated with SPMS—which can limit mobility, manual dexterity, and cognitive processing—and thereby reduce patients' ability or willingness to use the internet-based devices for medical consultations [23,41].

In Switzerland, telehealth expanded during the COVID-19 pandemic [8]. Remote consultations (mainly telephone and video) are reimbursed via existing tariff positions, and some basic insurance products use telemedicine "gatekeeper" models that require initial contact with a telemedicine provider. Despite this structural role, data on outpatient MS telehealth use remain limited, and our findings suggest that people living with MS do not yet view telehealth as an acceptable default for specialist care.

Contextualizing Our Findings Using the NASSS Framework

We further applied the NASSS framework to our findings to contextualize the results and to understand the complexity of implementing telehealth as a default in MS care [27]. Our review of the results identified four NASSS domains that are reflected by our study as follows:

1. Condition (MS complexity): MS is a heterogeneous disease with varying degrees of physical and cognitive impairment. Our finding of a negative association between SPMS and telehealth acceptance suggests that those with more severe or chronic forms of MS are less likely to accept telehealth by default, which is consistent with the literature [23,26,41].
2. Technology (telehealth use): the acceptance of telehealth as a default depends on prior digital literacy and technology exposure. In our cohort, only 20% had prior telemedicine experience, and over half expressed concerns about data security. These are moderating factors in digital health acceptance [44]. While global health bodies and the Swiss federal government promote digital health, acceptance may likely depend on how people living with MS engage with technology [1,32].
3. Value proposition (perceived benefits of telehealth)*:* over 70% of participants supported digital access to health data, but 80% declined telehealth as a default. Therefore, while most survey respondents are generally open toward provider-facing digital health tools in MS care, they seem more reluctant to actively engage with digital health technologies themselves. This finding could highlight a need to better communicate the potential benefits of telehealth to people living with MS [24,45]. While people with chronic diseases may accept certain aspects of digital health care [23], efforts are needed to better inform people living with MS on the value of telehealth to their health care.
4. Adopter system (integration into existing health care services)*:* one-third of participants who declined telehealth by default expressed the wish for support to use computer hardware or software. The presence of structured support systems is a critical factor in technology acceptance [7,44]. Addressing digital literacy gaps, easing the onboarding of persons with MS onto telehealth, and providing ongoing technical support and assistance could enhance the feasibility of telehealth as a default option for people living with MS [7,18].

While our analysis focused on 4 NASSS domains, we did not capture factors related to the organizational context, wider system, or long-term adaptation (Multimedia Appendix 4). This limits our understanding of how systemic factors—such as provider readiness, policy frameworks, and reimbursement models—may influence the acceptance of telehealth by default. These omissions are important, as they likely shape how people living with MS perceive and engage with telehealth beyond individual factors. Future research should prioritize collecting data on these broader domains to enable a more comprehensive use of the NASSS framework.

Strengths and Limitations

Our study has several limitations. First, given our recruitment methods and mode of survey delivery, our study population is not intended to be representative of the Swiss population of persons with MS at large (details provided in Iaquinto et al)[53]. Overall, our study sample seems to be older and more frequently of a PPMS or SPMS type. Therefore, we may have underestimated the general acceptability of telehealth (even after bias mitigation through reweighting our estimates based on our earlier, more inclusive digital health survey). However, the deviations of our (reweighted) samples from the Swiss population of persons with MS are unlikely to materially alter our main conclusion that the general openness to accept telehealth is presently quite low. Second, because the survey was administered online and focused on eHealth, it is possible that more digitally interested people living with MS were more likely to participate, which could in principle bias acceptance estimates upward, although the very low observed acceptance makes substantial overestimation unlikely.

Furthermore, since our study is a secondary data analysis, omitting some NASSS-relevant items, we were unable to assess certain NASSS domains, such as organizational factors, the wider health care system, and long-term sustainability. Also,

within the value proposition and adopter system domains, we were unable to consider the perspectives of health care providers and institutions. However, by assessing the perspective of people living with MS, our study offers important insights into the real-world acceptance of technology by the population of interest. Lastly, this study was conducted post-COVID-19 era, when the pressure to use telehealth is relatively low, and there is an increasing need for things to return to how they were before the pandemic, a possible reason for the low uptake.

However, our study has the following strengths. To our knowledge, this is the first study to specifically assess the acceptance of telehealth as a default in MS care. We also identified subgroups requiring targeted support to improve telehealth acceptance. Taken together, our study has identified some additional aspects that might pave the way toward an individualized telehealth setting suited for people with chronic diseases.

Implications and Recommendations

The low acceptance of telehealth as a default highlights the need for measures that address the barriers affecting acceptance. These measures may include targeted communication campaigns or having technical staff available to assist people living with MS with technical questions during onboarding. Our findings suggest that individuals who already engage with digital health tools for health management-related tasks, such as provider searches, are more receptive to telehealth. This suggests that furthering positive experiences with digital communication and provider interactions may also enhance the acceptance of telehealth. Furthermore, given that approximately one-third of participants required support with digital devices, implementing technical support and training programs could improve uptake. Policymakers and health care providers should prioritize these measures to ensure equitable digital health care access for people living with MS.

Clinically, our data argue against telehealth as a universal default model for MS care. Since participants with more severe disease were less likely to accept telehealth as the default modality. A more nuanced approach could involve triaging people living with MS according to disease severity, stability, and complexity: telehealth could be preferentially offered to those with stable, less severe disease for routine follow-up and counseling, whereas individuals with higher disability, new or complex symptoms, or substantial care coordination needs would continue to receive in-person visits by default. Such a stratified approach may optimize resource use while respecting individual preferences and clinical safety.

Conclusions

The acceptance and scaling of telehealth as a default is complex. Our study highlights the need for targeted implementation and testing of telehealth among a diverse cohort of people living with MS. These findings hint at potential considerations for the acceptance of telehealth as a default among people living with chronic diseases. Acceptance may be potentially improved if support is provided, and the value of these technologies is discussed with target patient groups. Such information should be accompanied by trust-building measures in data security and privacy, as well as offers for improving digital literacy for those in need. We conclude that acceptance of telehealth as a default needs to be targeted to the group of people living with chronic diseases most likely to benefit from it, and future research should explore provider perspectives and evaluate long-term strategies for integrating telehealth in MS care and chronic disease care.

Acknowledgments

The authors are grateful to Irene Rapold for her critical review of this manuscript. This manuscript was written without using generative artificial intelligence. The manuscript was drafted and reviewed in several rounds by all authors.

Funding

SNNE is funded by the Digital Society Initiative at the University of Zurich, Switzerland. The Swiss Multiple Sclerosis Registry is funded by the Swiss Multiple Sclerosis Society. These organizations were not involved in any study decisions.

Data Availability

The underlying data for this analysis are human research participant data, and in combination, are potentially identifiable. The data that supports the findings of this study are available upon reasonable request. Requests should be directed to the Swiss Multiple Sclerosis Registry at the University of Zurich; Epidemiology, Biostatistics & Prevention Institute; Hirschengraben 84, CH-8001 Zurich (ms-register@ebpi.uzh.ch).

Authors' Contributions

Conceptualization: SNNE, VVW
Data curation: SNNE, CPK, PC, CB, CG, CZ, AC, MP, VVW
Formal analysis: SNNE, VN, VVW
Funding acquisition: VVW
Methodology: SNNE, VVW
Supervision: VVW
Validation: SNNE, PD, VN, SI, EV-Y, CPK, PC, CB, CG, CZ, AC, MP, VVW
Writing – original draft: SNNE, PD, VN, VVW

Writing – review & editing: SNNE, PD, VN, SI, EV-Y, CPK, PC, CB, CG, CZ, AC, MP, VVW

Conflicts of Interest

AC has served on advisory boards for, and received funding for travel or speaker honoraria from Advisis, Alexion, Amgen, Biogen, BMS (Celgene), Dresden International Uni, Horizon Therapeutics, Janssen (J&J), Merck, Novartis, Roche, Sandoz, Sanofi-Genzyme, Teva, UCB, Uni Leipzig, Wiley, all for hospital research funds; and research support from Biogen, CSL Behring, Genzyme, Roche, UCB, and European Union and Swiss National Research Foundation. AC is the associate editor of the European Journal of Neurology and serves on the editorial board for Clinical and Translational Neuroscience and as topic editor for the Journal of International Medical Research.
CG: Ente Ospedaliero Cantonale (employer) received compensation for CG's speaking activities, consulting fees, or grants from Abbvie, Almirall, Biogen, Bristol Meyer Squibb, Lundbeck, Merck, Novartis, Sandoz, Sanofi, Teva Pharma, and Roche.
CZ: Ente Ospedaliero Cantonale (employer) received compensation for CZ's speaking activities, consulting fees, or grants from Abbvie, Alexion, Almirall, Biogen, Bristol Meyer Squibb, Eisai, Lilly, Lundbeck, Merck, Merz, Novartis, Organon, Pfizer, Sandoz, Sanofi, Teva Pharma, and Roche. CZ is a recipient of a grant for senior researchers provided by the Area Formazione accademica, Ricerca e Innovazione, EOC.
SNNE, PD, VN, SI, EV-Y, CK, PC, CB, MP, and VVW have no conflict of interest.

Multimedia Appendix 1
Swiss Multiple Sclerosis Registry (SMSR) participants in our digital survey.
[DOCX File, 18 KB - mhealth_v14i1e84447_app1.docx]

Multimedia Appendix 2
Use of technology and telemedicine experience.
[DOCX File, 21 KB - mhealth_v14i1e84447_app2.docx]

Multimedia Appendix 3
Univariable logistic regression analysis.
[DOCX File, 21 KB - mhealth_v14i1e84447_app3.docx]

Multimedia Appendix 4
Contextualizing our findings using the NASSS (Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability) framework.
[DOCX File, 20 KB - mhealth_v14i1e84447_app4.docx]

Checklist 1
STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist.
[PDF File, 116 KB - mhealth_v14i1e84447_app5.pdf]

References

1. Going digital for noncommunicable diseases: the case for action. : World Health Organization; 2024 URL: https://iris.who.int/bitstream/handle/10665/378478/9789240089921-eng.pdf [accessed 2025-12-23]
2. Ganguli I, Lim C, Daley N, Cutler D, Rosenthal M, Mehrotra A. Telemedicine adoption and low-value care use and spending among fee-for-service medicare beneficiaries. JAMA Intern Med 2025 Apr 1;185(4):440-449. [doi: 10.1001/jamainternmed.2024.8354] [Medline: 39992684]
3. Xiao Z, Han X. Evaluation of the effectiveness of telehealth chronic disease management system: systematic review and meta-analysis. J Med Internet Res 2023 Apr 27;25(1):e44256. [doi: 10.2196/44256] [Medline: 37103993]
4. Salisbury C, O'Cathain A, Thomas C, et al. Telehealth for patients at high risk of cardiovascular disease: pragmatic randomised controlled trial. BMJ 2016 Jun 1;353:i2647. [doi: 10.1136/bmj.i2647] [Medline: 27252245]
5. Hong YR, Xie Z, Nguyen OT, Turner K, Walker AF. Telehealth service use and quality of care among US adults with diabetes: a cross-sectional study of the 2022 health information national trends survey. BMJ Open 2024 Nov 3;14(10):e086418. [doi: 10.1136/bmjopen-2024-086418] [Medline: 39489536]
6. Shaffer KM, Turner KL, Siwik C, et al. Digital health and telehealth in cancer care: a scoping review of reviews. Lancet Digit Health 2023 May;5(5):e316-e327. [doi: 10.1016/S2589-7500(23)00049-3] [Medline: 37100545]
7. Learmonth YC, Galna B, Laslett LL, van der Mei I, Marck CH. Improving telehealth for persons with multiple sclerosis—a cross-sectional study from the Australian MS longitudinal study. Disabil Rehabil 2024 Oct;46(20):4755-4762. [doi: 10.1080/09638288.2023.2289594] [Medline: 38088346]
8. Nittas V, von Wyl V. COVID-19 and telehealth: a window of opportunity and its challenges. Swiss Med Wkly 2020 May 4;150(1920):w20284. [doi: 10.4414/smw.2020.20284] [Medline: 32400882]

9. Altmann P, Leutmezer F, Ponleitner M, et al. Remote visits for people with multiple sclerosis during the COVID-19 pandemic in Austria: the TELE MS randomized controlled trial. Digit Health 2022;8:20552076221112154. [doi: 10.1177/20552076221112154] [Medline: 35847524]
10. Yeroushalmi S, Maloni H, Costello K, Wallin MT. Telemedicine and multiple sclerosis: a comprehensive literature review. J Telemed Telecare 2020;26(7-8):400-413. [doi: 10.1177/1357633X19840097] [Medline: 31042118]
11. Ehde DM, Arewasikporn A, Alschuler KN, Hughes AJ, Turner AP. Moderators of treatment outcomes after telehealth self-management and education in adults with multiple sclerosis: a secondary analysis of a randomized controlled trial. Arch Phys Med Rehabil 2018 Jul;99(7):1265-1272. [doi: 10.1016/j.apmr.2017.12.012] [Medline: 29337024]
12. Turner AP, Wallin MT, Sloan A, et al. Clinical management of multiple sclerosis through home telehealth monitoring: results of a pilot project. Int J MS Care 2013;15(1):8-14. [doi: 10.7224/1537-2073.2012-012] [Medline: 24453757]
13. Gromisch ES, Turner AP, Neto LO, Haselkorn JK, Raskin SA. Improving prospective memory in persons with multiple sclerosis via telehealth: a randomized feasibility study. Mult Scler Relat Disord 2024 Aug;88:105718. [doi: 10.1016/j.msard.2024.105718] [Medline: 38878624]
14. Kever A, Aguerre IM, Vargas W, et al. Feasibility trial of a telehealth support group intervention to reduce anxiety in multiple sclerosis. Clin Rehabil 2022 Oct;36(10):1305-1313. [doi: 10.1177/02692155221107077] [Medline: 35673256]
15. Leavitt VM, Riley CS, De Jager PL, Bloom S. eSupport: Feasibility trial of telehealth support group participation to reduce loneliness in multiple sclerosis. Mult Scler 2020 Nov;26(13):1797-1800. [doi: 10.1177/1352458519884241] [Medline: 31668134]
16. Van Denend T, Mathiowetz V, Preissner K, et al. Adapting the multiple sclerosis functional composite for telehealth administration using videoconference delivery: methodological considerations and interrater reliability. Arch Rehabil Res Clin Transl 2024 Jun;6(2):100337. [doi: 10.1016/j.arrct.2024.100337] [Medline: 39006110]
17. Khan F, Amatya B, Kesselring J, Galea M, editor. Telerehabilitation for persons with multiple sclerosis. Cochrane Database Syst Rev 2015 Apr 9;2015(4):CD010508. [doi: 10.1002/14651858.CD010508.pub2] [Medline: 25854331]
18. Landi D, Ponzano M, Nicoletti CG, et al. Patient's point of view on the use of telemedicine in multiple sclerosis: a web-based survey. Neurol Sci 2022 Feb;43(2):1197-1205. [doi: 10.1007/s10072-021-05398-6] [Medline: 34283343]
19. Roth EG, Minden SL, Maloni HW, Miles ZJ, Wallin MT. A qualitative, multiperspective inquiry of multiple sclerosis telemedicine in the United States. Int J MS Care 2022;24(6):275-281. [doi: 10.7224/1537-2073.2021-117] [Medline: 36545645]
20. Uribe Guajardo MG, Baillie A, Louie E, et al. The evaluation of the role of technology in the pathways to comorbidity care implementation project to improve management of comorbid substance use and mental disorders. J Multimorb Comorb 2022;12:26335565221096977. [doi: 10.1177/26335565221096977] [Medline: 35586033]
21. Jafri RF, Jafri NF, Sayeed M, Musthafa M, Riyas S. Telehealth in multiple sclerosis: fostering quality of life. Mult Scler Relat Disord 2023 Dec;80:105280. [doi: 10.1016/j.msard.2023.105280]
22. Leal Neto O, Von Wyl V. Digital transformation of public health for noncommunicable diseases: narrative viewpoint of challenges and opportunities. JMIR Public Health Surveill 2024 Jan 25;10:e49575. [doi: 10.2196/49575] [Medline: 38271097]
23. Marrie RA, Kosowan L, Cutter G, Fox R, Salter A. Disparities in telehealth care in multiple sclerosis. Neur Clin Pract 2022 Jun;12(3):223-233. [doi: 10.1212/CPJ.0000000000001167]
24. Blut M, Chong AYL, Tsigna Z, et al. Meta-analysis of the unified theory of acceptance and use of technology (UTAUT): challenging its validity and charting a research agenda in the Red Ocean. J Assoc Inf Syst 2022 Jan;23(1):13-95. [doi: 10.17705/1jais.00719]
25. Salisbury C, Thomas C, O'Cathain A, et al. Telehealth in chronic disease: mixed-methods study to develop the TECH conceptual model for intervention design and evaluation. BMJ Open 2015 Feb 6;5(2):e006448. [doi: 10.1136/bmjopen-2014-006448] [Medline: 25659890]
26. Shin HD, Hamovitch E, Gatov E, et al. The NASSS (Non-Adoption, Abandonment, Scale-Up, Spread and Sustainability) framework use over time: a scoping review. PLOS Digit Health 2025 Mar;4(3):e0000418. [doi: 10.1371/journal.pdig.0000418] [Medline: 40096260]
27. Greenhalgh T, Wherton J, Papoutsi C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 1;19(11):e367. [doi: 10.2196/jmir.8775] [Medline: 29092808]
28. Greenhalgh T, Rosen R. Remote by default general practice: must we, should we, dare we? Br J Gen Pract 2021 Apr;71(705):149-150. [doi: 10.3399/bjgp21X715313] [Medline: 33771790]
29. Li V, Roos I, Monif M, et al. Impact of telehealth on health care in a multiple sclerosis outpatient clinic during the COVID-19 pandemic. Mult Scler Relat Disord 2022 Jul;63:103913. [doi: 10.1016/j.msard.2022.103913] [Medline: 35661564]
30. Barin L, Kaufmann M, Salmen A, et al. Patterns of care for multiple sclerosis in a setting of universal care access: a cross-sectional study. Mult Scler Relat Disord 2019 Feb;28:17-25. [doi: 10.1016/j.msard.2018.11.033] [Medline: 30530118]
31. Stanikić M, Twomey E, Puhan MA, et al. Experiences of persons with multiple sclerosis with the Covid-19 vaccination: a cross-sectional study of the Swiss Multiple Sclerosis Registry. Mult Scler Relat Disord 2023 Jun;74:104707. [doi: 10.1016/j.msard.2023.104707] [Medline: 37068368]

32. Zimmer A. Transformation numérique du système de santé: gare aux chantiers [Article in French]. Bull Med Suisses 2022;103(48):26-27. [doi: 10.4414/bms.2022.21300]
33. Xiang XM, Bernard J. Telehealth in multiple sclerosis clinical care and research. Curr Neurol Neurosci Rep 2021 Feb 28;21(4):14. [doi: 10.1007/s11910-021-01103-4] [Medline: 33646409]
34. Greenhalgh T, Abimbola S. The NASSS framework—a synthesis of multiple theories of technology implementation. Stud Health Technol Inform 2019 Jul 30;263:193-204. [doi: 10.3233/SHTI190123] [Medline: 31411163]
35. Steinemann N, Kuhle J, Calabrese P, et al. The Swiss Multiple Sclerosis Registry (SMSR): study protocol of a participatory, nationwide registry to promote epidemiological and patient-centered MS research. BMC Neurol 2018 Aug 13;18(1):111. [doi: 10.1186/s12883-018-1118-0] [Medline: 30103695]
36. Puhan MA, Steinemann N, Kamm CP, et al. A digitally facilitated citizen-science driven approach accelerates participant recruitment and increases study population diversity. Swiss Med Wkly 2018;148(1920):w14623. [doi: 10.4414/smw.2018.14623] [Medline: 29767828]
37. Nittas V, Zecca C, Kamm CP, Kuhle J, Chan A, von Wyl V. Digital health for chronic disease management: an exploratory method to investigating technology adoption potential. PLOS ONE 2023;18(4):e0284477. [doi: 10.1371/journal.pone.0284477] [Medline: 37053272]
38. Abimbola S, Patel B, Peiris D, et al. The NASSS framework for ex post theorisation of technology-supported change in healthcare: worked example of the TORPEDO programme. BMC Med 2019 Dec 30;17(1):233. [doi: 10.1186/s12916-019-1463-x] [Medline: 31888718]
39. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
40. Chen MH, Goverover Y, Botticello A, DeLuca J, Genova HM. Healthcare disruptions and use of telehealth services among people with multiple sclerosis during the COVID-19 pandemic. Arch Phys Med Rehabil 2022 Jul;103(7):1379-1386. [doi: 10.1016/j.apmr.2021.12.028] [Medline: 35093328]
41. Özden F, Özkeskin M, Ekici E, Tümtürk İ, Ekmekci Ö, Yüceyar N. Opinions, satisfaction and expectations of individuals with multiple sclerosis about telerehabilitation services. Clin Neurol Neurosurg 2024 Feb;237:108162. [doi: 10.1016/j.clineuro.2024.108162] [Medline: 38325037]
42. Ward M, Goldman MD. Epidemiology and pathophysiology of multiple sclerosis. Continuum (Minneap Minn) 2022 Aug 1;28(4):988-1005. [doi: 10.1212/CON.0000000000001136] [Medline: 35938654]
43. Lee EC, Grigorescu V, Enogieru I, et al. Updated national survey trends in telehealth utilization and modality (2021-2022). : Office of the Assistant Secretary for Planning and Evaluation (ASPE); 2022 Mar URL: https://www.ncbi.nlm.nih.gov/books/NBK604393/ [accessed 2025-12-23]
44. Venkatesh V, Thong J, Xu X, University of Arkansas, Hong Kong University of Science and Technology, The Hong Kong Polytechnic University. Unified theory of acceptance and use of technology: a synthesis and the road ahead. JAIS 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
45. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inf Manage 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
46. Raja M, Bjerkan J, Kymre IG, Galvin KT, Uhrenfeldt L. Telehealth and digital developments in society that persons 75 years and older in European countries have been part of: a scoping review. BMC Health Serv Res 2021 Oct 26;21(1):1157. [doi: 10.1186/s12913-021-07154-0] [Medline: 34696789]
47. Bhatia R, Gilliam E, Aliberti G, et al. Older adults' perspectives on primary care telemedicine during the COVID-19 pandemic. J Am Geriatr Soc 2022 Dec;70(12):3480-3492. [doi: 10.1111/jgs.18035] [Medline: 36169152]
48. Nguyen OT, Alishahi Tabriz A, Huo J, Hanna K, Shea CM, Turner K. Impact of asynchronous electronic communication-based visits on clinical outcomes and health care delivery: systematic review. J Med Internet Res 2021 May 5;23(5):e27531. [doi: 10.2196/27531] [Medline: 33843592]
49. Sadeghi N, Eelen P, Nagels G, et al. Innovating care in multiple sclerosis: feasibility of synchronous internet-based teleconsultation for longitudinal clinical monitoring. J Pers Med 2022 Mar 10;12(3):433. [doi: 10.3390/jpm12030433] [Medline: 35330433]
50. Vesinurm M, Maunula A, Olli P, et al. Effects of a digital care pathway for multiple sclerosis: observational study. JMIR Hum Factors 2024 Aug 7;11(1):e51872. [doi: 10.2196/51872] [Medline: 39110966]
51. D'Haeseleer M, Eelen P, Sadeghi N, D'Hooghe MB, Van Schependom J, Nagels G. Feasibility of real time internet-based teleconsultation in patients with multiple sclerosis: interventional pilot study. J Med Internet Res 2020 Aug 13;22(8):e18178. [doi: 10.2196/18178] [Medline: 32447274]
52. Cimperman M, Makovec Brenčič M, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
53. Iaquinto S, Chan A, Manjaly ZM, et al. Rising prevalence of multiple sclerosis in Switzerland: results from the Swiss Multiple Sclerosis Registry. Neuroepidemiology 2025;59(6):623-632. [doi: 10.1159/000542632] [Medline: 39557017]

Abbreviations

aOR: adjusted odds ratio
CIS: clinically isolated syndrome
CVD: cardiovascular disease
HCP: health care professional
MS: multiple sclerosis
NASSS: Non-Adoption, Abandonment, Scale-Up, Spread, and Sustainability
PPMS: primary progressive multiple sclerosis
RRMS: relapsing-remitting multiple sclerosis
SMSR: Swiss Multiple Sclerosis Registry
SPMS: secondary progressive multiple sclerosis
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by L Buis; submitted 19.Sep.2025; peer-reviewed by ES Yoon, R Haase; revised version received 16.Dec.2025; accepted 16.Dec.2025; published 23.Jan.2026.

Please cite as:
Ekongefeyin SNN, Daniore P, Nittas V, Iaquinto S, Vallejo-Yagüe E, Kamm CP, Calabrese P, Baum C, Gobbi C, Zecca C, Chan A, Puhan M, von Wyl V
Acceptability of Telehealth as the Default Modality for Multiple Sclerosis Care in Switzerland: Cross-Sectional Study
JMIR Mhealth Uhealth 2026;14:e84447
URL: https://mhealth.jmir.org/2026/1/e84447
doi:10.2196/84447

Review

Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Minyue Sun[1], MSN; Fuyou Tang[2], MSN; Luo min[1], BS; Shiyu Wen[1], BS; Shuang Wang[3], BS; Huiping Jiang[3], BS

[1]Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, China

[2]Mianyang Key Laboratory of Anesthesia and Neuroregulation, Department of Anesthesiology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

[3]Department of Neurology, Mianyang Central Hospital, School of Medicine, University of Electronic Science and Technology of China, Mianyang, Sichuan, China

Corresponding Author:
Luo min, BS
Department of Neurology
Mianyang Central Hospital, School of Medicine
University of Electronic Science and Technology of China
Mianyang
China
Phone: 86 18324126071
Email: 410244386@qq.com

Abstract

Background: The global integration of telehealth into the management of Parkinson disease (PD) addresses critical gaps in health care access, especially for patients with limited mobility in underserved regions. Despite accelerated adoption during the COVID-19 pandemic, evidence regarding telehealth's multidimensional efficacy remains inconsistent. Previous meta-analyses reported conflicting outcomes for quality of life (QOL), motor symptoms, and neuropsychiatric comorbidities.

Objective: This study aimed to quantitatively synthesize the effects of telehealth interventions across six core PD domains: (1) QOL, (2) depression, (3) anxiety, (4) motor symptoms, (5) activities of daily living (ADL), and (6) cognition.

Methods: PubMed, Embase, Cochrane Library, Scopus, and Web of Science were systematically searched until June 21, 2024. In adherence to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, English-language randomized controlled trials evaluating telehealth interventions for PD were included. Study quality was assessed using the Cochrane Risk of Bias tool. A dual analytical approach using random-effects models was applied to address heterogeneity. Studies reporting a single effect size were analyzed using the Hartung-Knapp-Sidik-Jonkman correction. Studies with multiple dependent effect sizes were analyzed using a 3-level random-effects meta-analysis with t-distribution inference, accounting for sampling, within-study, and between-study variance. Effect sizes were expressed as standardized mean differences (SMD) with 95% CIs. Heterogeneity was quantified using the τ^2; prediction intervals were not calculated due to the limited number of studies. Prespecified subgroup analyses examined intervention types (digital vs traditional telehealth) and follow-up durations. Sensitivity analyses and assessments for small-study effects (multilevel Egger tests, funnel plots) were conducted.

Results: A total of 15 randomized controlled trials (765 participants) demonstrated significant telehealth benefits: QOL significantly improved on the Medical Outcomes Study 36-Item Short Form Health Survey and Brunnsviken Brief Quality of Life Scale (SMD 0.39, 95% CI 0.06-0.72; *P*=.03), with marginal improvement on the Parkinson Disease Questionnaire-8 (SMD –0.42, 95% CI –0.88 to 0.03; *P*=.07). Telephone-based interventions outperformed digital approaches (*P*=.002). Depression symptoms were significantly reduced (SMD –0.64, 95% CI –0.93 to 0.34; *P*<.001), particularly with traditional telehealth (*P*<.001). Anxiety also decreased significantly (SMD –0.64, 95% CI –0.92 to 0.35; *P*=.003) with negligible heterogeneity (I^2=0%). Motor symptoms improved (SMD –0.46, 95% CI –0.69 to 0.24; *P*=.001), and ADL showed substantial impairment reduction (SMD –0.79, 95% CI –1.04 to –0.54; *P*=.002). Cognition was significantly enhanced (SMD 1.12, 95% CI 0.03 to 2.20; *P*=.045) though with moderate heterogeneity (I^2=52.3%) and significant publication bias (*P*<.001). Follow-up duration did not significantly moderate effects.

Conclusions: Telehealth interventions significantly enhance multiple PD domains, with traditional (telephone/tablet-based) approaches demonstrating particular advantages for QOL and depression. Digital interventions showed more limited efficacy. These findings support telehealth as a multifaceted management tool for PD, although cognition outcomes require further investigation.

Trial Registration: PROSPERO CRD42024520169; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024520169

(JMIR Mhealth Uhealth 2026;14:e70994) doi:10.2196/70994

KEYWORDS

depression; meta-analysis; motor symptoms; Parkinson disease; quality of life; systematic review; telehealth intervention

Introduction

Parkinson disease (PD), currently the second-most common neurodegenerative disorder after Alzheimer disease, affects approximately 1% of individuals aged 55 years or older, typically manifesting around the age of 60 years [1]. The incidence of PD increases with advancing age, particularly among individuals older than 60 years [1]. Currently, PD affects nearly 6 million people globally [2]. By 2030, China is expected to have approximately 4.94 million patients with PD, representing nearly half of global cases [3]. This demographic shift could significantly strain national economies and health care systems [3].

Patients with PD experience a range of motor and nonmotor symptoms. Motor symptoms include tremors, rigidity, bradykinesia, postural instability, and gait freezing [4]. Nonmotor symptoms encompass cognitive decline, pain, fatigue, psychiatric conditions such as depression and anxiety, and sleep disturbances [5,6]. These symptoms substantially reduce patients' quality of life (QOL) and increase psychological stress and physical demands on caregivers [7]. Traditionally, management and evaluation for patients with PD require visits to outpatient clinics or hospitals for advanced diagnostics and assessments [8]. However, many patients encounter significant barriers to accessing such facilities, including limited mobility, fear of falling, depressive symptoms, fatigue, and time constraints. These barriers can worsen their symptoms, delay treatment, lead to potentially life-threatening complications, and increase overall disease burden [9]. Consequently, a growing number of researchers advocate telehealth as a means to improve PD diagnosis, treatment, and rehabilitation. Telehealth aims to overcome challenges associated with in-person consultations and geographical disparities in health care resources, thereby reducing delays in treatment, decreasing morbidity and mortality, and improving QOL among patients with PD [10].

Telehealth uses digital information and communication technologies to connect patients with health care providers and deliver medical services [11]. These technologies include internet-connected desktop computers, tablets, smartphones, and wearable devices [12,13]. The COVID-19 pandemic significantly accelerated the adoption of telehealth among patients with PD, improving health care accessibility. Research supports the practicality of telehealth and underscores its perceived effectiveness by patients with PD and neurologists [10,14-16]. Neurologists can independently conduct comprehensive assessments using the Movement Disorders Society-Unified Parkinson's Disease Rating Scale Part III (MDS-UPDRS-III), enabling more accurate evaluations in patients' usual settings rather than clinical environments. This method provides a more accurate reflection of the patients' actual condition [10,17-19]. Advantages of telehealth, such as time and cost efficiency, have resulted in high patient satisfaction [14-16,20-22]. Additionally, telehealth facilitates virtual monitoring for rehabilitation, psychotherapy, and advanced PD treatments [23-25]. It promotes interdisciplinary collaboration and provides education and training opportunities for physicians and health care workers in developing regions, overcoming geographic, travel, and financial barriers [10].

Numerous studies have confirmed the practicality and effectiveness of telehealth in managing patients with PD; yet, findings regarding QOL, anxiety, depression, motor function, activities of daily living (ADL), and cognitive function have been variable [8,26,27]. A meta-analysis by Chen et al [26] in 2020 demonstrated that telehealth interventions effectively reduced motor symptoms compared with traditional care. However, these interventions showed no substantial improvements in QOL, depression, cognitive functions, or balance abilities, contrasting with more recent findings [28-31]. A systematic review by Leon-Salas et al [8] published in 2023 indicated limited and inconclusive data regarding telehealth services for patients with PD. Following the COVID-19 pandemic, an influx of new studies necessitates an updated synthesis and analysis to clarify the effects of telehealth interventions on patients with PD. Additionally, a 2024 systematic review by Federico et al [27] reported telerehabilitation outcomes comparable to in-person therapy for patients with PD. However, the review focused solely on telerehabilitation for various neurological disorders without extensively evaluating broader telehealth effects in patients with PD.

Thus, the objective of this research was to systematically compile and analyze the most recent data on telehealth intervention effects in patients with PD.

Methods

Overview

This systematic review and meta-analysis was conducted in strict accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) [32]. To ensure transparency and reproducibility, the study protocol was registered with PROSPERO (registration number CRD42024520169).

Inclusion and Exclusion Criteria

The selection criteria followed the PICOS framework: (1) Population: individuals officially diagnosed with PD; (2) Interventions: telehealth or telemedicine interventions delivered via telephone, internet, or other digital communication technologies [11]; (3) Comparison/Control: studies with an experimental group compared to control groups using standard care, routine care, conventional care, or waitlist control; (4) Outcomes: assessment of intervention effects on overall health or specific behavioral and psychological symptoms associated with Parkinsonism; (5) Study type: only randomized controlled trials (RCTs).

The exclusion criteria were as follows: (1) studies published in non-English languages, (2) incomplete studies, such as research protocols or ongoing studies, (3) studies where interventions were exclusively telehealth-based without a comparison group, or where no telehealth intervention was applied, (4) studies lacking sufficient details on relevant outcome measures, and (5) studies without adequate statistical data for analysis. No publication date limitations were applied.

Search Strategy

Two researchers (MS and FT) systematically searched 5 English-language databases (PubMed, Embase, Cochrane Library, Scopus, and Web of Science) from database inception until June 21, 2024. The comprehensive search strategy combined subject headings and keywords related to three main topics: (1) PD, (2) telehealth, and (3) RCTs. There were no limits on publication status or dates. The detailed search methods are provided in Multimedia Appendix 2. Additionally, references of included studies were manually reviewed.

Study Selection and Data Extraction

The reference management tool EndNote X9 (Clarivate Analytics) was used for data management. After removing duplicates, 2 reviewers (MS and FT) independently screened titles and abstracts based on predefined inclusion and exclusion criteria. Potentially relevant articles underwent full-text assessment to determine eligibility. Any discrepancies were resolved through discussion, and when consensus was unattainable, a third reviewer (Luomin) was consulted for a final decision.

Data extraction used a specifically designed form. Extracted information included authors' names, publication dates, country, study design, sample size, demographic and clinical characteristics (average age, gender distribution, and disease attributes), type and duration of interventions and control groups, outcome measures, and key study findings.

Risk of Bias

Two researchers (MS and FT) independently assessed the methodological quality and potential bias of included studies using the Cochrane Collaboration's Risk of Bias tool [33]. Disagreements were resolved through discussion with a third reviewer (Luomin) to achieve consensus. The tool systematically evaluates several dimensions of bias, including random sequence generation, allocation concealment, blinding of participants and personnel, handling of incomplete outcome data, and selective outcome reporting. Each study was assessed for potential selection, performance, detection, attrition, and reporting biases. Risks in each domain were categorized as low (unlikely to significantly affect results), high (likely to significantly undermine confidence), or unclear.

Statistical Analysis

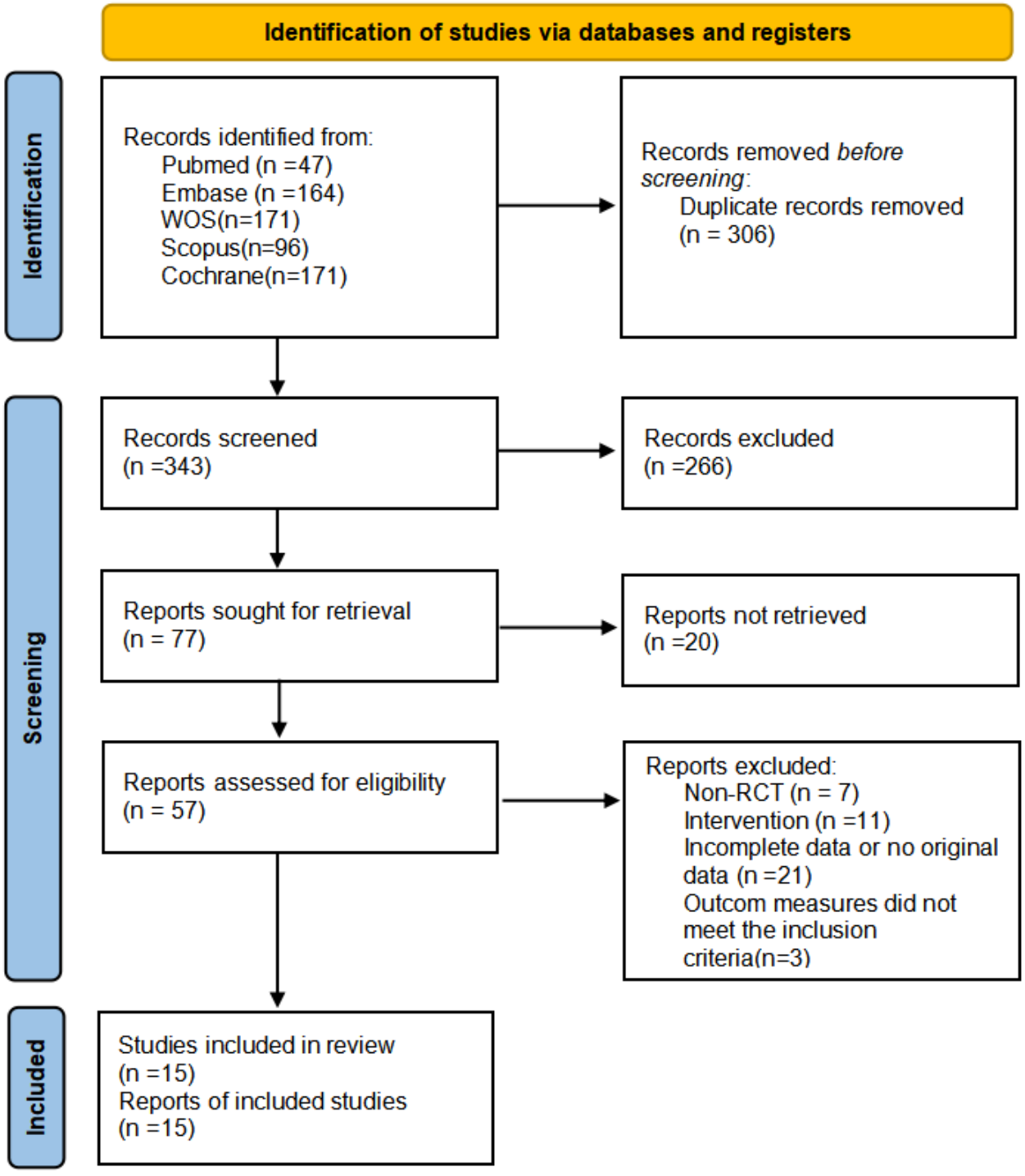
All meta-analyses were conducted using random-effects models, based on the conceptual assumption that true effects vary across studies due to inherent differences in populations, interventions, and settings, rather than on statistical metrics such as I^2. A dual analytical approach incorporating the Hartung-Knapp-Sidik-Jonkman (HKSJ) adjustment was used to obtain more accurate and conservative interval estimates: for studies providing a single effect size, the HKSJ method was used to calculate 95% CIs; for studies with multiple dependent effect sizes, a 3-level random-effects meta-analysis was performed in R software using the *metafor* package (version 4.5.1; R Foundation for Statistical Computing) using t-distribution–based inference. This model accounts for sampling variance (level 1), within-study variance (level 2), and between-study variance (level 3), with t-distribution inference being analogous to the HKSJ correction for multilevel models. Heterogeneity was quantified with the τ^2 statistic. Although 95% prediction intervals were initially planned to illustrate the expected range of true effects in similar future studies, they were not calculated due to the limited number of studies for each outcome (all $k<10$), as such intervals are unreliable with small samples. The I^2 statistic is reported for descriptive purposes only, acknowledging its limited pragmatic use in conveying the magnitude of true effect variation across settings. Prespecified subgroup analyses examined intervention type (digital vs traditional telehealth) and follow-up duration. Sensitivity analyses were conducted using leave-one-out elimination. Small-study effects (for which publication bias is one possible explanation) were assessed with funnel plots and multilevel Egger tests. Statistical significance was set at 2-tailed $P<.05$. Complete analysis code is provided in Multimedia Appendix 3.

Results

Study Selection

Figure 1 shows the PRISMA flowchart detailing the study selection process for this systematic review and meta-analysis. The initial electronic database search yielded 649 records. After removing duplicates, 343 articles remained. Following title and abstract screening, 266 articles were excluded, leaving 77 articles eligible for full-text review. Among these, 20 articles (conference abstracts, reviews, or research protocols) were unavailable in full, leaving 57 articles for detailed evaluation. Ultimately, 15 articles met the inclusion criteria and were included in the final analysis.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart detailing the identification and selection of randomized controlled trials (RCTs) on telehealth interventions for people with Parkinson disease. WOS: Web of Science.



Study Characteristics

This systematic review included 15 RCTs [20,28-31,34-43] with a total of 765 participants. The studies were conducted across various countries: 2 in Spain, 6 in the United States, 1 in Japan (Tokyo), 3 in Italy, 1 in Sweden, 1 in Brazil, and 1 in Australia. These studies were published between 2016 and 2024. All participants were diagnosed with PD and aged 18 years or older. Sample sizes ranged from 9 to 49 participants. All trials used telehealth technologies, including telehealth systems, telephones, applications, and tablets. Follow-up durations ranged from 1 to 12 months post intervention. Detailed study characteristics are provided in Table 1.

Table 1. Characteristics of randomized controlled trials evaluating telehealth interventions for people with Parkinson disease (N=15 studies, 2016-2024).

Author, year	Country	Sample size, T/C[a]	Age (years), T/C[b], mean (SD)	Participant	Diagnostic criteria	Telehealth technology	Follow-up time	Outcomes (with measure scales)
Cubo et al, 2017 [34]	Spain	17/18	66.44 (7.09)/ 66.05 (9.76)	PD[c]	Medical records	Telehealth system (Kinesia system included a tablet software app, a wireless finger-worn motion sensor unit, and automated web-based symptom reporting.)	12 months	PD severity (UPDRS[d], parts I-IV), the severity of nonmotor symptoms (Non-Motor Quest), QOL[e] (EQ-5D), depression and anxiety (HADS[f]), caregiver burden (Zarit Burden)
Del Pino et al, 2023 [35]	Spain	10/10	64.5 (7.9)/ 69.1 (3.5)	PD	UK Brain Bank	Telemedicine system (vCare system: a virtual training platform created based on an intelligent ICT[g] environment for rehabilitation of neurological and cardiac diseases related to aging)	4 months	QOL (Euroqol 5D), cognitive general status (MoCA[h]), PD severity (UPDRS, parts I-IV), functional disability (H&Y[i]), England ADL[j]
Dobkin et al, 2020 [36]	United States	37/35	65.62 (9.76)/ 64.80 (9.62)	PD and caregivers	National Institute of Neurological Disorders and Stroke research criteria	Telephone (using the telephone for CBT[k])	End of treatment/6 months	Depression (HAM-D[l], BDI[m]), anxiety (HAM-A[n]), QOL (SF-36[o])
Dobkin et al, 2021 [30]	United States	45/45	67.27 (7.79)/66.42 (9.51)	PD	Medical records	Telephone (using the telephone for CBT)	End of treatment/6 months	Depression (HAM-D, BDI), anxiety (HAM-A), QOL (SF-36)
Duffley et al, 2021 [37]	United States	23/19	65.0 (10.9)/64.1 (10.0)	PD and caregivers	Medical records	Telephone (using the telephone for health guidance and follow-up)	6 months	Motor symptoms (UPDRS, part III), QOL (PDQ-39[p]), caregiver burden (MCSI[q]),
Eldemir et al, 2023 [38]	Turkey	15/15	57.87 (9.79)/61.40 (7.29)	PD	UK Brain Bank	Telephone (carrying out rehabilitation training courses through telephone videoconferences)	1.5 months	ADL (UPDRS, part II), motor symptoms (UPDRS, part III), QOL (PDQ-8)
Ellis et al, 2018 [39]	United States	23/21	64.8 (8.5)/63.3 (10.6)	PD	UK Brain Bank	App (Wellpepper app: this health app provides a detailed exercise plan, including what, how, when, and where to perform exercises. Push notifications motivate users to complete their exercise and walking programs. A physical therapist remotely adjusts the regimen based on user progress. Visual progress tracking helps users monitor their performance throughout the program.)	12 months	QOL (PDQ-39), walking capacity (6-MWT[r])

Author, year	Country	Sample size, T/C[a]	Age (years), T/C[b], mean (SD)	Participant	Diagnostic criteria	Telehealth technology	Follow-up time	Outcomes (with measure scales)
Gandolfi et al, 2017 [40]	Italy	36/34	67.45 (7.18)/69.84 (9.41)	PD	UK Brain Bank	App (TeleWii-Lab: contains the Nintendo Wii console for motion control input, the Wii Fit game system, and the balance board. A laptop computer connected to a high-definition webcam was used to establish real-time remote video communication between the rehabilitation unit and the patient's residence via Skype software.)	End of treatment/1 month	QOL (PDQ-8, walking capacity (10-MWT[s]), balance (BBS[t], ABC[u])
GoffrEdo et al, 2023 [41]	Italy	49/48	67.8 (6.6)/68.2 (5.8)	PD	UK Brain Bank	Tablet (carry out motor and cognitive rehabilitation training through the VRRS[v] tablet system based on nonimmersive virtual reality.)	1.5-2.5 months	Walking capacity (6-MWT, TUG[w]), motor symptoms (UPDRS, part III), balance (mini-Balance Evaluation)
Heldman et al, 2017 [42]	United States	9/9	65.2 (10.1)/68.6 (10.2)	PD	Medical records	Telephone (conducting videoconferences or providing telephone guidance for implementing interventions via telephone)	7 months	PD severity (UPDRS, part I-IV), QOL (PDQ-39), the Patient Assessment of Chronic Illness Care
Kraepelien et al, 2020 [29]	Sweden	38/39	65.9 (8.5)/66.1 (9.8)	PD	Medical records	Telephone (using the telephone for CBT)	1.25 months/2.5 months	Depression and anxiety (HADS), QOL (PDQ-8, BBQ[x])
Maggio et al, 2024 [31]	Italy	T1: 12; T2: 12/10	T1: 59.7 (9.7); T2: 63.8 (8.3)/66.8 (6.5)	PD	Medical records	App (nonimmersive VR[y] app: NeuroNation Brain Training by Synaptikon GMBH, Berlin, offering science-based mental training to enhance various cognitive abilities with personalized data reports, and Train Your Brain by Grove FX, focusing on specific skills like concentration, spatial thinking, and reasoning. Additionally, there's a social cognitive app known as Sims Mobile.)	1.5 months/3 months	Cognitive general status (MMSE[z], MoCa), depression and anxiety (HAM-D)
Pastana Ramos et al, 2023 [28]	Brazil	8/11	60.7 (17.04)/58.6 (8.15)	PD	UK Brain Bank	Tablet\telephone (individualized telerehabilitation sessions were conducted using a tablet or mobile phone through videoconferencing and verbal guidance.)	1 month/2 months	Walking capacity (TUG, 5STS[aa]), balance (ABC), QOL (PDQ-8), motor symptoms (MDS-UPDRS[ab], part III)
Theodoros et al, 2016 [43]	Australia	15/16	71.62 (7.77)/72.86 (9.99)	PD	Medical records	Telemedicine system (eHAB: a mobile multimedia telerehabilitation system that offers real-time videoconferencing and transmits treatment data to the user's computer as images and texts. It also records high-definition live video and audio.)	1 month	QOL (PDQ-39)

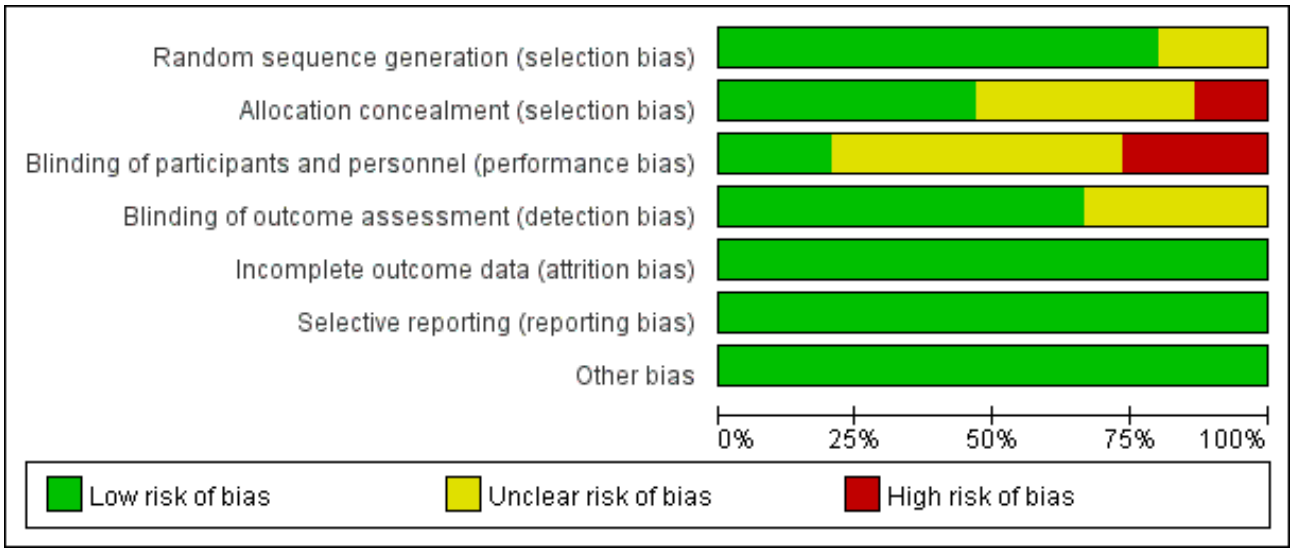
Author, year	Country	Sample size, T/C[a]	Age (years), T/C[b], mean (SD)	Participant	Diagnostic criteria	Telehealth technology	Follow-up time	Outcomes (with measure scales)
Wilkinson et a, 2016 [20]	United States	Arm 1: 26/24; Arm 2: 18/18	Arm 1: 76.1 (8.4)/76.1 (7.9); Arm 2: 67.2 (9.8)/70.9 (8.4)	PD	UK Brain Bank	Telemedicine system (a global health care telemedicine specialist cart and Cisco webcam provide real-time high-definition audio-visual connectivity between patients and health care providers.)	6 months/12 months	PD severity (Arm 1: UPDRS, part I-IV; Arm 2: H&Y), QOL (PDQ-8), depression and anxiety (GDS[ac])

[a]T/C: sample size in treatment group/control group.
[b]T/C: mean (SD) values in treatment group/control group.
[c]PD: Parkinson disease.
[d]UPDRS: Unified Parkinson's Disease Rating Scale.
[e]QOL: quality of life.
[f]HADS: Hospital Anxiety and Depression Scale.
[g]ICT: digital information and communication technologies.
[h]MoCA: Montreal Cognitive Assessment.
[i]H&Y: Hoehn and Yahr Scale.
[j]ADL: activities of daily living.
[k]CBT: cognitive behavioral therapy.
[l]HAM-D: Hamilton Depression Rating Scale.
[m]BDI: Beck Depression Inventory.
[n]HAM-A: Hamilton Anxiety Rating Scale.
[o]SF-36: Medical Outcomes Study 36-Item Short Form Health Survey.
[p]PDQ: Parkinson Disease Questionnaire.
[q]MCSI: Michigan Consumer Sentiment Index.
[r]6-MWT: 6-minute walk test.
[s]10-MWT: 10-minute walk test.
[t]BBS: Berg Balance Scale.
[u]ABC: Activities-specific Balance Confidence scale.
[v]VRRS: Virtual Router Redundancy Service.
[w]TUG: Timed Up and Go Test.
[x]BBQ: Brunnsviken Brief Quality of Life Scale.
[y]VR: virtual reality.
[z]MMSE: Mini-Mental State Examination.
[aa]5STS: 5 Times Sit-to-Stand Test.
[ab]MDS-UPDRS: Movement Disorders Society-Unified Parkinson's Disease Rating Scale.
[ac]GDS: Geriatric Depression Scale.

Risk of Bias

The methodological quality and potential biases of the included studies were assessed using the Cochrane Risk of Bias tool [33]. Figure 2 presents detailed findings. All 15 studies were confirmed as randomized, and 12 studies [20,28,30,31,34,36-41,43] clearly described methods of random sequence generation. Only 7 studies [28-30,36,38,39,41] detailed allocation concealment methods, categorizing them as having a low risk of selection bias. Regarding the blinding of participants and intervention providers, only 3 studies [36,39,41] reported adequate blinding procedures and thus had a low risk of performance bias. In contrast, 4 studies [28,37,40,43] without blinding were classified as having a high risk of performance bias. The remaining 8 studies [20,29-31,34,35,38,42] lacked sufficient information and were classified as unclear. Blinding of outcome assessors was reported in 10 studies [28,30,36-43], indicating a low risk of detection bias, while 5 studies [20,29,31,34,35] did not report this information. All studies provided complete outcome data, indicating low attrition bias. Assessment for other potential biases generally indicated low risk. Funnel plot analysis (Multimedia Appendix 4) revealed no significant publication bias among included studies.

Figure 2. Risk of bias assessment (using the Cochrane Risk of Bias tool) for the included randomized controlled trials evaluating telehealth interventions for Parkinson disease.



Meta-Analysis

Quality of Life

A total of 11 studies [20,28-30,36-40,42,43] evaluated QOL following telehealth interventions, using 5 different measurement tools. Overall, 9 studies used Parkinson Disease Questionnaire (PDQ) scales (PDQ-39 [37,39,42,43] and PDQ-8 [20,28,29,38,40]), with higher scores indicating worse QOL. Additionally, 2 studies used the Medical Outcomes Study 36-Item Short Form Health Survey (SF-36) [30,36], and 1 study incorporated both PDQ-8 and Brunnsviken Brief Quality of Life Scale (BBQ) [29], with higher scores reflecting better QOL. Due to measurement variability, a standardized mean difference (SMD)–based meta-analysis was conducted following Cochrane handbook guidelines.

A 3-level random-effects meta-analysis of 8 studies (13 effect sizes) assessing QOL using PDQ scales revealed a marginal trend toward improvement after telehealth interventions (SMD –0.42, 95% CI –0.88 to 0.03; P=.07). Heterogeneity was present (τ^2=0.258; I^2=56.2%; Q_{12}=27.37; P=.007), with between-study variance accounting for 65.5% of total variability. Prediction intervals were not calculated due to the limited number of studies (k<10; Table 2; Figures S1-S3 in Multimedia Appendix 4). Subgroup analysis demonstrated significant moderation by intervention type (test of moderators=9.84; P=.002). Telephone-based interventions significantly improved QOL (SMD –0.83, 95% CI –1.22 to –0.44; P<.001), whereas digital interventions had minimal effect (SMD –0.05, 95% CI –0.58 to 0.49; P=.86). The between-group difference was 0.78 (95% CI 0.29-1.27; P=.002) points. Follow-up duration did not significantly moderate outcomes (categorical: P=.58; continuous: β=.01/month; P=.83; Table 3; Figures S4-S7 in Multimedia Appendix 4). Sensitivity analysis identified substantial influence from Ramos et al [28], whose exclusion reduced the effect magnitude by 52.2%. Multilevel Egger test suggested possible small-study effects (P=.02; Figures S8-S9 in Multimedia Appendix 4).

Table 2. Results of 3-level random-effects meta-analyses for primary outcomes in patients with Parkinson disease receiving telehealth interventions.

Outcome	*k* [a]	*#ES* [b]	Mg[c]	95% CI	*P* value	σ^2_{level2}[d]	σ^2_{level3}[e]	Variance level 1 (%)	Variance level 2 (%)	Variance level 3 (%)
QOL[f,g]	8	13	–0.42	–0.88 to 0.034	.07	0.000	0.258	34.5	0	65.5
QOL[h]	3	5	0.39	0.06 to 0.72	.03	0.019	0.000	72.3	27.7	0
Anxiety	3	5	–0.64	–0.92 to –0.35	.003	0.000	0.000	100	0	0
Cognitive function	2	9	1.12	0.03 to 2.20	.045	0.192	0.275	33.9	27.2	38.7
Depression	5	17	–0.64	–0.93 to –0.34	<.001	0.016	0.062	55.4	8.9	35.6
Motor symptoms	7	11	–0.46	–0.69 to –0.24	<.001	0.000	0.000	100	0	0

[a]*k*=number of studies.

[b]*#ES*=number of effect sizes.

[c]Mg=mean effect size (g).

[d]σ^2_{level2}=variance between effect sizes extracted from the same study.

[e]σ^2_{level3}=variance between studies.

[f]QOL: quality of life.

[g]The 3-level meta-analysis of 8 studies (13 effect sizes) assessing QOL with Parkinson Disease Questionnaire.

[h]The 3-level meta-analysis of 3 studies (5 effect sizes) assessing QOL with Medical Outcomes Study 36-Item Short Form Health Survey and Brunnsviken Brief Quality of Life Scale.

Table 3. Moderator analyses examining the effects of intervention type and follow-up duration on primary outcomes in the meta-analysis.

	k [a]	*#ES* [b]	B_0/g [c]	t_0 [d]	B_1 [e]	t_1 [f]	*F* test (*df*) [g]
QOL[h,i]							
Intervention type	8	13	–0.83	–4.179[j]	0.78	3.137[j]	9.842[j] (1, 11)
Follow duration[k]	8	13	–0.53	–1.852	0.26	0.550	0.302 (1, 11)
Follow duration[l]	8	13	–0.46	–0.167	0.01	0.219	0.048 (1, 11)
QOL[m]							
Follow duration[k]	3	5	0.73	3.033[j]	–0.42	–0.157	2.470 (1, 3)
Follow duration[l]	3	5	0.69	3.369[n]	–0.06	–1.713	2.937 (1, 3)
Anxiety							
Follow duration[k]	3	5	–0.51	–2.206[o]	–0.16	–0.605	0.366 (1, 3)
Follow duration[l]	3	5	–0.66	–3.179[j]	0.004	0.124	0.015 (1, 3)
Cognitive function							
Follow duration[k]	2	9	1.39	2.876[j]	–0.31	–0.654	0.427 (1, 7)
Follow duration[l]	2	9	2.07	3.125[j]	–0.30	–0.186	1.407 (1, 7)
Depression							
Intervention type	5	17	–0.84	–9.933[n]	0.53	3.416[n]	11.669[n] (1, 15)
Follow duration 1[k]	5	17	–0.54	–1.979[o]	–0.13	–0.419	0.175 (1, 15)
Follow duration 2[l]	5	17	—0.87	–4.798[n]	0.04	1.860	0.346 (1, 15)
Motor symptoms							
Intervention type	7	11	—0.54	—3.542[n]	0.13	0.660	0.436 (1, 9)
Follow duration[k]	7	11	–0.58	–3.624[n]	0.20	0.949	0.901 (1, 9)
Follow duration[l]	7	11	–0.55	–3.037[j]	0.015	0.595	0.354 (1, 9)

[a]k=number of independent studies.

[b]*#ES*=number of effect sizes.

[c]B_0/mean g=intercept/mean effect size (g).

[d]t_0=*t* value for mean g.

[e]B_1=estimated regression coefficient.

[f]t_1=*t* value for regression coefficient.

[g]Omnibus *F* test.

[h]QOL: quality of life.

[i]The 3-level meta-analysis of 8 studies (13 effect sizes) assessing QOL with Parkinson Disease Questionnaire.

[j]*P*<.01.

[k]Results of analysis based on categorical variables.

[l]Results of analysis based on continuous variables.

[m]The 3-level meta-analysis of 3 studies (5 effect sizes) assessing QOL with the Medical Outcomes Study 36-Item Short Form Health Survey and Brunnsviken Brief Quality of Life Scale.

[n]*P*<.001.

[o]*P*<.05.

For 3 studies assessing QOL using SF-36 and BBQ (5 effect sizes), the 3-level random-effects meta-analysis indicated significant improvement after telehealth interventions (SMD 0.39, 95% CI 0.06-0.72; *P*=.03). Minimal heterogeneity was observed (τ^2=0.020; I^2=25.9%; Q_4=5.40; *P*=.25), with within-study variance contributing 27.7% of total variability. Prediction intervals were not calculated (*k*<10; Table 2; Figures S11-S12 in Multimedia Appendix 4). Subgroup analyses indicated greater effects in short-term follow-ups (<3 months: SMD 0.73, 95% CI 0.26-1.19) compared with longer-term

follow-ups (≥3 months: SMD 0.31, 95% CI –0.40 to 0.43; β=–.42; *P*=.12). A marginal negative correlation between effect magnitude and follow-up duration was observed (β=–.06/month; *P*=.09; Table 3; Figures S13-S15 in Multimedia Appendix 4). Sensitivity analysis confirmed robustness; no single-study exclusion altered significance substantially (maximum change: +33.2% when excluding Dobkin et al [36]). The Egger test indicated no significant small-study effects (*P*=.28; Figures S16-S17 in Multimedia Appendix 4).

Depression

A total of 5 studies [20,29-31,36] evaluated depression levels in patients with PD following telehealth interventions, using 4 distinct assessment tools: Hamilton Depression Rating Scale (HAM-D) [30,31,36], Beck Depression Inventory (BDI) [30,36], Hospital Anxiety and Depression Scale-Depression (HADS-D) [29], and Geriatric Depression Scale [20]. Higher scores on these scales indicate increased depression severity. Two studies used both the HAM-D and BDI to assess depression. A 3-level random-effects meta-analysis (5 studies, 17 dependent effect sizes) with t-distribution–based inference revealed a significant reduction in depression symptoms after telehealth interventions (SMD –0.64, 95% CI –0.93 to –0.34; *P*<.001). Moderate heterogeneity was observed (I^2=45.9%; Q_{16}=29.60; *P*=.02). Variance component analysis indicated that between-study differences accounted for 35.6% of total variability (τ^2=0.062), within-study variability explained 8.9% (τ^2=0.016), and sampling error contributed 55.4% (Table 2; Figures S17-S19 in Multimedia Appendix 4). Prediction intervals were not calculated because the number of studies was below the recommended threshold (*k*<10). Subgroup analysis showed significantly greater improvements with traditional telehealth (SMD –0.84, 95% CI –1.00 to –0.67; *P*<.001) compared with digital interventions (SMD –0.31, 95% CI –0.54 to –0.08; *P*=.008; β=.53; *P*<.001). Follow-up duration was not significant in categorical analysis (β=–.13; *P*=.68), but continuous analysis showed a marginal positive association (β=.04/month; *P*=.06; Table 3; Figures S20-S23 in Multimedia Appendix 4). Sensitivity analyses confirmed robustness; effects remained significant after excluding each study (SMD range –0.55 to –0.79). The Egger test indicated no significant small-study effects (*P*=.82; Figures S24-S25 in Multimedia Appendix 4).

Anxiety

A total of 3 studies [29,30,36] assessed anxiety levels in patients with PD using 2 measurement scales: Hamilton Anxiety Rating Scale (HAM-A) [30,36] and Hospital Anxiety and Depression Scale-Anxiety (HADS-A) [29]. Higher scores represent greater anxiety severity. A 3-level random-effects meta-analysis (3 studies, 5 dependent effect sizes) with t-distribution–based inference indicated significant anxiety reduction following telehealth interventions (SMD –0.64, 95% CI –0.92 to –0.35; *P*=.003). Negligible heterogeneity was observed (I^2=0%; Q_4=0.90; *P*=.92), with variance component analysis indicating that sampling error explained all variability (τ^2=0 for both inter- and intrastudy variance; Table 2; Figures S26-S28 in Multimedia Appendix 4). Prediction intervals were not calculated due to the limited number of studies (*k*<10). All studies used telephone-based interventions, thus subgroup analyses examined follow-up durations. Follow-up duration was not significantly associated with outcomes in categorical (β=–.16, 95% CI –0.66 to 0.35; *P*=.55) or continuous analyses (β=.004/month, 95% CI –0.06 to 0.07; *P*=.90; Table 3; Figures S29-S31 in Multimedia Appendix 4). Sensitivity analyses confirmed robustness, with effects remaining significant upon exclusion of individual studies (SMD range –0.67 to –0.60), and maximum deviation of –5.4% when excluding Dobkin et al [30]. Funnel plots were symmetrical, and the Egger test revealed no significant small-study effects (*P*=.68; Figures S32-S33 in Multimedia Appendix 4).

Motor Symptoms

A total of 7 studies [20,28,34,35,38,41,42] used the MDS-UPDRS-III to evaluate motor symptoms in patients with PD post telehealth interventions, with higher scores indicating greater severity. A 3-level random-effects meta-analysis (7 studies, 11 dependent effect sizes) with t-distribution–based inference demonstrated significant improvements in motor symptoms after telehealth interventions (SMD –0.46, 95% CI –0.69 to –0.24; *P*=.001). Heterogeneity among studies was negligible (I^2=0%; Q_{10}=7.85; *P*=.64). Variance component analysis showed that sampling error accounted for all variability, with negligible between-study (τ^2=0) and within-study variance (τ^2=0; Table 2; Figures S34-S36 in Multimedia Appendix 4). Prediction intervals were not calculated because the number of studies was below the recommended threshold (*k*<10). Subgroup analysis by intervention type revealed no significant difference between digital and other telehealth interventions (β=.13, 95% CI –0.26 to 0.53; *P*=.51). Follow-up duration did not significantly moderate outcomes in categorical (β=.20, 95% CI –0.21 to 0.60; *P*=.34) or continuous analyses (β=.02/month, 95% CI –0.03 to 0.06; *P*=.55; Table 3; Figures S37-S39 in Multimedia Appendix 4). Sensitivity analyses confirmed robustness, with significant effects maintained after each study's exclusion (SMD range –0.52 to –0.42), and the largest change being 11.5% when excluding Wilkinson et al [20]. Funnel plots showed symmetry, and the Egger test indicated no significant small-study effects (*P*=.88; Figures S40-S41 in Multimedia Appendix 4).

Activities of Daily Living

A total of 4 studies [34,35,38,42] evaluated the impact of telehealth interventions on daily activities in patients with PD using the MDS-UPDRS-II scale. Higher scores represent greater impairment. A random-effects meta-analysis with HKSJ correction demonstrated that telehealth interventions significantly reduced impairment in daily activities compared with controls (SMD –0.79, 95% HKSJ-adjusted CI –1.04 to –0.54; *P*=.002). Heterogeneity was negligible (τ^2=0.000; I^2=0.0%; Q_3=0.43; *P*=.93). Prediction intervals were not calculated because the number of studies was below the recommended threshold (*k*<10). Sensitivity analysis (leave-one-out) confirmed the robustness of the findings. The largest change in effect size occurred after excluding Del Pino et al [35] (–7.9% change; SMD range across exclusions –0.83 to –0.73). Egger test indicated no significant small-study effects

(intercept=–1.35; *P*=.31; Figure S42 in Multimedia Appendix 4).

Cognition

A total of 2 studies [31,35] examined cognitive outcomes in patients with PD after telehealth interventions using the Montreal Cognitive Assessment (MoCA) and Mini-Mental State Examination (MMSE), where higher scores indicate better cognitive function. One study used both the MoCA and MMSE scales to comprehensively assess cognition. A 3-level random-effects meta-analysis (2 studies, 9 dependent effect sizes) with t-distribution–based inference indicated significant cognitive improvement following telehealth interventions (SMD 1.12, 95% CI 0.03-2.20; *P*=.045). Moderate heterogeneity was detected (I^2=52.3%; Q_8=16.77; *P*=.03). Variance component analysis showed that between-study differences accounted for 38.9% (τ^2=0.275) of total variability, within-study differences accounted for 27.2% (τ^2=0.192), and sampling error explained 33.9% (Table 2; Figures S43-S45 in Multimedia Appendix 4). Prediction intervals were not calculated due to the limited number of studies (*k*<10). A follow-up duration did not significantly moderate outcomes in categorical (β=–.31, 95% CI –1.25 to 0.62; *P*=.51) or continuous analysis (β=–.30/month, 95% CI –0.79 to 0.20; *P*=.24; Table 3; Figures S46-S48 in Multimedia Appendix 4). The Egger test indicated significant small-study effects (intercept=15.17; *P*<.001). Sensitivity analysis was not feasible due to the limited number of studies (Figure S49 in Multimedia Appendix 4).

Discussion

Overview

This systematic review evaluated the effects of telehealth interventions on the QOL and associated health outcomes in patients with PD. Findings demonstrated that telehealth interventions significantly enhanced various dimensions of patient well-being, including QOL, depressive symptoms, anxiety levels, motor function, ADL, and cognitive abilities. These results differ from earlier reviews and meta-analyses [8,26,27]. The discrepancies might be due to accelerated advancements in telehealth and remote neurology after the epidemic, alongside improvements in telehealth service quality and increased research volume [22,44]. Although the effectiveness of telehealth interventions appears promising, additional studies are needed to establish more conclusive evidence.

Effectiveness of Telehealth Interventions on QOL in Patients With PD

Our meta-analysis identified complex patterns regarding telehealth's effect on QOL in patients with PD. Interventions assessed by the SF-36/BBQ demonstrated a significant improvement, whereas those assessed using PDQ scales indicated only marginal benefit. This difference likely stems from the fundamental distinctions between scales: PDQ scales specifically measure PD-related deficits, whereas SF-36/BBQ assess general well-being [45,46]. Notably, telephone-based interventions substantially improved PDQ-based QOL, while digital interventions showed negligible effects. This distinction explains inconsistencies in prior meta-analyses [1,26,27] that did not account for intervention modality. Significant heterogeneity in PDQ analyses was primarily attributed to between-study differences, suggesting methodological variations. Sensitivity analysis highlighted substantial influence from Ramos et al [28], and potential small-study effects were noted. These factors necessitate cautious interpretation of PDQ outcomes. Conversely, SF-36/BBQ analyses exhibited minimal heterogeneity and robust sensitivity. Differences in results may also reflect variations in intervention type, therapeutic intensity, assessment timing, methodological quality, and sample sizes [27]. Follow-up duration did not moderate effects, contradicting assumptions that longer interventions yield superior outcomes. Instead, the intervention modality emerged as the critical moderator, highlighting the importance of direct human interaction in managing PD-specific QOL concerns. This aligns with telehealth's capacity for dynamic therapeutic engagement [47,48], particularly beneficial for isolated patients [10], although our findings indicate that these advantages depend on modality.

Telehealth provides more dynamic, immersive methods for treatment, education, and counseling compared to traditional medical approaches, enhancing patient engagement and interaction [47,48]. Such enhancements assist patients with PD and their families in comprehensively understanding and managing disease-related challenges, thus promoting independence, motivation for self-care, and improved life quality [49]. Patients with PD require consistent engagement with health care teams for effective management of disease progression and treatment complexity [8]. A primary advantage of telehealth lies in serving patients in isolated or underserved areas, addressing health care provider shortages, and offering timely, high-quality care to improve patient outcomes [10]. Additionally, economic burdens and logistical difficulties substantially reduce patients with PD's QOL [50,51]. By reducing health care–related costs and travel demands, telehealth can expand home-based medical services, further enhancing life quality for patients with PD [10,22].

Effectiveness of Telehealth Interventions on Depression in Patients With PD

Our meta-analysis demonstrated a significant antidepressant effect of telehealth interventions. Traditional approaches (telephone or cognitive behavioral therapy [CBT]–based) showed nearly 3 times the efficacy compared to digital interventions. This advantage aligns with neurobiological evidence linking depression in PD to dysfunction in serotonergic pathways and frontostriatal circuits [52], suggesting human-mediated therapies more effectively modulate emotional processing compared to automated digital tools. This finding notably diverges from earlier studies, which primarily emphasized motor symptoms and physical rehabilitation, often neglecting depressive symptoms [10,27]. The distinct focus of 3 specific studies [29,30,36] included in this meta-analysis may explain this difference. These studies emphasized cognitive and behavioral aspects of patient care, integrating telehealth with CBT, a method recognized for effectively reducing depression levels [53,54]. Moderate heterogeneity mainly resulted from

between-study methodological differences, likely reflecting variations in measurement tools and intervention protocols. Contrary to expectations, categorical follow-up duration showed no significant moderating effect, although continuous analysis revealed a marginal positive association. This result suggests sustained engagement—particularly through telephone-based CBT [53,54]—might progressively reinforce neuroplastic changes in emotion-regulation networks. These findings reconcile previous contradictions in the literature [1,26,27]. Whereas earlier reviews primarily targeted motor symptoms, our analysis confirms telehealth's antidepressant benefits when including behavioral interventions tailored to PD-related psychopathology.

Our findings align with those of Dou et al [55], indicating that telehealth interventions (tele-CBT and telerehabilitation training) significantly improve depressive symptoms in patients with PD. Feasibility has been verified internationally; for example, a cross-sectional study in Brazil showed effective telehealth implementation in resource-limited settings with high patient satisfaction [56,57]. Regarding specific methods, tele-motor training significantly enhanced patients' motor function and indirectly alleviated depressive symptoms [55]. In contrast, tele-CBT directly targeted depressive and anxiety symptoms, showing greater effectiveness compared to other teleinterventions [31]. Teleconsultation had relatively limited efficacy in alleviating depressive symptoms but significantly improved access to medical resources [58]. From a neurobiological standpoint, PD and depression share common pathological mechanisms, including gut microbiota dysregulation, neuroinflammation, and reward-processing dysfunction [52]. Telehealth, especially through behavioral interventions such as CBT, may modulate these pathological processes and consequently alleviate depressive symptoms [59]. However, existing evidence suggests telehealth's effectiveness may be weaker for chronic, nonepisodic mental disorders (eg, depression in PD) compared to primary depression [60]. To optimize telehealth potential, future research should investigate long-term outcomes, standardization of techniques, and cybersecurity considerations [61,62]. In summary, telehealth effectively reduces depressive symptoms in patients with PD, especially via tele-CBT, which overcomes geographical barriers and improves treatment accessibility. Nevertheless, individualized plans and sustained follow-up are necessary to achieve optimal therapeutic outcomes.

Effectiveness of Telehealth Interventions on Anxiety in Patients With PD

Our meta-analysis demonstrated robust anxiolytic effects of telephone-based telehealth interventions, with remarkable consistency across studies. This homogeneity suggests that telephone-delivered CBT provides a reliably standardized approach for managing PD-related anxiety. Notably, these benefits remained stable irrespective of follow-up duration, indicating sustained therapeutic effects without attenuation over 3-9 months. These findings resolve previous contradictions [14,26,27] by demonstrating that structured tele-CBT can effectively address PD-specific anxiety mechanisms, including fear-avoidance cycles and "off"-period distress resistant to conventional treatments. The negligible heterogeneity, with variance entirely attributable to sampling error, likely reflects 3 factors. First, interventions used standardized CBT protocols targeting PD-specific anxiety mechanisms such as hypervigilance toward motor fluctuations. Second, the uniform application of validated and sensitive scales (HAM-A/HADS-A) ensured measurement precision. Third, telephone delivery strengthened therapeutic alliances through real-time emotional interaction absent in purely digital interfaces. The integration of standardized protocols, precise assessments, and person-centered delivery resulted in methodological consistency across studies.

Although considerable evidence supports telehealth for anxiety in patients with PD, its exact mechanism and broader applicability require further investigation. Previous studies [63,64] showed comparable efficacy of telehealth and face-to-face interventions in reducing anxiety, depression, and stress scores, alongside improved heart rate variability. Anxiety reductions persisted long-term after telehealth interventions, confirming their noninferiority to in-person care. This effectiveness largely stems from multimodal interventions; for example, remote CBT overcomes movement-related barriers and, combined with exercise and biomarker monitoring, allows personalized care beneficial to underserved populations [55,64,65]. However, some research highlights intervention heterogeneity. A small study [66] indicated that telephone CBT effectively alleviated depression but not anxiety symptoms in patients with PD, suggesting anxiety may require more tailored strategies. Our analysis, in contrast, supports the long-term feasibility, effectiveness, and durability of telephone CBT effects. Earlier discrepancies might stem from small sample sizes or limitations of measurement tools. Although the revised Parkinson Anxiety Scale improved cultural adaptability, general scales (eg, HADS) may underestimate actual effectiveness due to limited sensitivity [29,67]. Future research should expand sample sizes, develop PD-specific anxiety interventions, and integrate multidimensional biomarker monitoring to improve telehealth precision and applicability.

Effectiveness of Telehealth Interventions on Motor Symptoms of Patients With PD

Telehealth interventions significantly improved motor symptoms in patients with PD. This refined estimate may reflect advancements in methodological rigor involving multilevel analyses that account for independent effect sizes, an approach not consistently used in previous meta-analyses [1,26]. Due to the standardized use of MDS-UPDRS-III assessments and similar intensities of interventions, we observed remarkably low heterogeneity among studies. Notably, digital and traditional telehealth approaches demonstrated comparable effectiveness, indicating that essential motor rehabilitation components, such as amplitude training and balance exercises, effectively translated across different treatment platforms. The temporal stability of benefits further supported telehealth as a sustainable management option, with sensitivity analyses confirming robustness to study exclusion.

A primary therapeutic objective in PD involves improving motor symptoms, wherein treatment adjustments frequently depend on accurate motor assessments [68]. The telehealth framework

enables improved and timely interactions between patients and health care providers compared to traditional face-to-face consultations, allowing for more individualized rehabilitation strategies tailored specifically to patients with movement disorders [69]. Multiple studies have confirmed the significant impact of telehealth interventions in alleviating motor symptoms in patients with PD. For instance, structured telerehabilitation programs, such as the Lee Silverman Voice Treatment BIG rehabilitation method, have effectively enhanced motor function, alleviated nonmotor symptoms, and improved the QOL for patients with PD [70]. Compared to teleconsultations alone, tele-motor interventions demonstrate superior efficacy in motor function improvement [55]. From a neuromechanism perspective, cueing techniques activate the motor cortex, thereby enhancing the stability of motor output, which provides scientific justification for using cue-based strategies in telerehabilitation [71]. Moreover, telerehabilitation is particularly suitable for patients with restricted mobility or those residing in medically underserved regions. Real-time video guidance ensures continuous rehabilitation training, effectively overcoming geographical limitations. Its safety and potential effectiveness in improving balance and functional activities have been confirmed by existing research [28,41,72]. Telehealth facilitates comprehensive monitoring of treatment effects through standardized scales (such as MDS-UPDRS) for assessing motor symptoms, combined with evaluation of nonmotor symptoms and QOL questionnaires [73,74]. Additionally, tele-motor interventions based on live-streaming have been proven feasible and safe, demonstrating high patient adherence (eg, twice a week) and thus confirming their practical use for continuous management of motor symptoms in PD [75]. Therefore, telehealth effectively enhances motor functions in patients with PD, offering advantages in personalized program design, activation of neural plasticity, and overcoming limitations in medical resource availability. With ongoing advancements in assessment instruments and technological integration, telehealth is anticipated to further improve long-term intervention outcomes.

Effectiveness of Telehealth Interventions on ADL in Patients With PD

The results of this study showed that telehealth interventions significantly improved ADL among patients with PD. This result aligns with previous research, reinforcing that remote health care interventions significantly enhance both ADL performance and motor symptoms in individuals with PD. Our analysis suggests that the significant improvements in ADL resulting from telehealth are due to multidimensional intervention strategies addressing the core symptoms of PD.

Relevant studies have shown that structured remote rehabilitation programs, delivered through real-time video instruction, enhance functional mobility and directly improve basic ADL tasks such as walking and dressing [57]. Simultaneously, high-intensity remote exercise interventions reduce motor sluggishness and freezing of gait, indirectly enhancing instrumental ADLs, such as complex daily activities like shopping and meal preparation [76]. The simultaneous improvements observed in ADLs and motor symptoms share clear pathophysiological connections; enhanced motor functions directly alleviate limitations in physical activity, enabling patients to execute daily routines more effectively [77]. Additionally, remote CBT improves executive functions, mitigating motor-related restrictions on complex ADL performance [77,78]. Telehealth frameworks achieve these synergistic effects by integrating 3 primary components: real-time video supervision ensures adherence to exercise regimens; home-based cognitive training modules restructure the prefrontal-limbic circuitry; and wearable sensors provide immediate feedback regarding movement quality [57,79,80]. Therefore, telehealth interventions positively and synergistically influence both motor symptoms and ADL performance in patients with PD. Future research should focus on optimizing intervention strategies and integrating motor and ADL training components comprehensively to further enhance the overall QOL for patients.

Effectiveness of Telehealth Interventions on Cognition in Patients With PD

Preliminary evidence shows that telehealth interventions may enhance cognitive function in patients with PD; however, these findings should be interpreted cautiously. Although statistically significant, effect sizes exhibited substantial variability, ranging from negligible to considerable clinical improvement. This observed heterogeneity primarily stems from methodological differences among studies, potentially reflecting (1) the use of varied cognitive assessments (MoCA vs MMSE), each with differing sensitivities to PD-specific cognitive deficits, and (2) distinct intervention protocols within the limited scope of available evidence. Additionally, significant publication bias and insufficient data for sensitivity analyses further limit definitive conclusions.

Overall, the efficacy of telehealth interventions for enhancing cognitive functions in patients with PD has been established. These interventions significantly improve cognitive status, particularly executive functions and memory, as well as emotional and behavioral disorders, consequently enhancing the QOL for both patients and caregivers [78,81]. Among specific intervention methods, computer-assisted cognitive training has shown potential benefits for patients with PD accompanied by mild cognitive impairment, with feasibility confirmed for home-based training modalities [82,83]. Moreover, remote virtual reality applications (telehealth virtual reality) have shown promising results for improving cognitive task performance [31]. A recent network meta-analysis further supports the beneficial effects of remote interventions, including remote cognitive training, on cognition and other nonmotor symptoms [55]. The primary advantages of telehealth interventions include high accessibility (especially beneficial for patients with limited mobility or those residing in remote areas) and flexibility, with the patient's cognitive reserve potentially enhancing treatment effect [84]. However, considerable heterogeneity exists within current evidence, aligning with our findings. This heterogeneity is largely attributed to variations in study design, inconsistencies in cognitive assessment tools, and diverse responses among patient subtypes [55,82,85]. In addition, the efficacy of telehealth interventions differed across cognitive domains. Therefore, these findings should be considered exploratory and interpreted

cautiously. In conclusion, telehealth represents a promising cognitive management approach for PD with substantial potential; nevertheless, implementation barriers must be considered, strategies tailored to individual patient needs, and larger standardized trials conducted to further substantiate effectiveness.

Strengths and Limitations

This systematic review benefits from a rigorous methodological approach, using a meta-analysis grounded in RCTs and strictly adhering to established guidelines for systematic reviews. All analyses were conducted using random-effects models based on conceptual considerations, with HKSJ or t-distribution–based corrections applied to provide more accurate and conservative CIs. This significantly enhances the credibility of the findings. Furthermore, the review assesses the impact of telehealth interventions not only on QOL but also on multiple health-related domains such as depression, anxiety, motor function, ADL, and cognitive function in patients with PD, rather than restricting its focus solely to treatment modalities.

Nonetheless, several limitations of this review should be acknowledged. First, the limited number of RCTs included in the analysis may constrain the generalizability of these conclusions to broader populations. Second, due to the small number of studies per outcome (all $k<10$), prediction intervals were not calculated, which limits the interpretation of how the true effect may vary across different settings. Third, funnel plots and Egger tests were used to assess small-study effects, but these methods have reduced accuracy when fewer than 10 studies are analyzed per outcome, and they do not specifically measure publication bias. Finally, although we applied multilevel modeling to account for dependent effect sizes, residual heterogeneity, and variations in intervention protocols may still influence the results. Therefore, additional RCTs must be incorporated into future research to enhance the robustness and reliability of the findings.

Implications for Practice

Global disparities in medical resource distribution present substantial challenges to health care service advancement. This issue is particularly pronounced in neurological care, where specialist availability is limited, notably in suburban and rural regions. Consequently, many individuals with PD struggle to receive continuous medical support, resulting in significant declines in their QOL as the disease progresses. This situation places considerable strain not only on patients and their families but also on societal resources. The emergence of telehealth, however, offers a promising solution by providing innovative avenues for managing and treating PD. Telehealth has the potential to bridge existing gaps, enabling patients with PD who previously had limited or no access to receive essential health care services.

Implications for Further Research

PD exerts substantial impacts on public health, prompting significant attention from the health care community toward preventative, diagnostic, and therapeutic strategies. As an innovative product of rapid technological advancement, telehealth represents a cost-effective, real-time, and secure platform for collecting patient data, significantly facilitating the diagnosis, monitoring, and rehabilitation of PD. Nonetheless, the efficacy of telehealth requires further validation through comprehensive and rigorous RCTs. Future research should not only evaluate functional recovery, cognitive enhancement, and health-related QOL but also examine aspects such as cost-effectiveness, patient satisfaction, and digital health literacy among older adults. Such investigations will facilitate more informed decisions and optimal tailoring of telehealth interventions for patients with PD.

Conclusion

Telehealth interventions have demonstrated the potential to significantly enhance various aspects of life among patients with PD, including alleviating symptoms of depression and anxiety, improving motor function, facilitating ADL, and enhancing cognitive performance. Despite these encouraging findings, there remains an urgent need for meticulously designed, large-scale RCTs to comprehensively evaluate telehealth's effectiveness across the full spectrum of PD management.

Acknowledgments

We would like to express our heartfelt thanks to all those involved in this study.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: MS, FT, Luomin
Data curation: MS, FT
Formal analysis: MS, FT
Investigation: MS, FT
Methodology: FT, Luomin
Project administration: Luomin
Software: MS
Validation: S Wang, S Wen

Visualization: HJ
Writing - original draft: MS
Writing - review & editing: MS, FT, Luomin, S Wen, S Wang, HJ

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA 2020 checklist.
[DOCX File , 29 KB - mhealth_v14i1e70994_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOC File , 45 KB - mhealth_v14i1e70994_app2.doc]

Multimedia Appendix 3
R language code.
[DOCX File , 25 KB - mhealth_v14i1e70994_app3.docx]

Multimedia Appendix 4
Supplementary figures.
[DOC File , 6938 KB - mhealth_v14i1e70994_app4.doc]

References

1. Lee A, Gilbert RM. Epidemiology of parkinson disease. Neurol Clin 2016;34(4):955-965. [doi: 10.1016/j.ncl.2016.06.012] [Medline: 27720003]
2. Bisaglia M, Filograna R, Beltramini M, Bubacco L. Are dopamine derivatives implicated in the pathogenesis of Parkinson's disease? Ageing Res Rev 2014;13:107-114. [doi: 10.1016/j.arr.2013.12.009] [Medline: 24389159]
3. Li G, Ma J, Cui S, He Y, Xiao Q, Liu J, et al. Parkinson's disease in China: a forty-year growing track of bedside work. Transl Neurodegener 2019;8:22 [FREE Full text] [doi: 10.1186/s40035-019-0162-z] [Medline: 31384434]
4. Engelender S, Isacson O. The threshold theory for Parkinson's disease. Trends Neurosci 2017;40(1):4-14. [doi: 10.1016/j.tins.2016.10.008] [Medline: 27894611]
5. Sekirina TP, Voronkova TL, Tsutsu'lkovskaia MI, Abramova LI. Disorders of interleukin 2 biosynthesis by peripheral blood lymphocytes in patients with schizophrenia. Zh Nevropatol Psikhiatr Im S S Korsakova 1989;89(5):95-97. [Medline: 2789461]
6. Przedborski S. The two-century journey of Parkinson disease research. Nat Rev Neurosci 2017;18(4):251-259. [doi: 10.1038/nrn.2017.25] [Medline: 28303016]
7. Macchi ZA, Koljack CE, Miyasaki JM, Katz M, Galifianakis N, Prizer LP, et al. Patient and caregiver characteristics associated with caregiver burden in Parkinson's disease: a palliative care approach. Ann Palliat Med 2020;9(Suppl 1):S24-S33 [FREE Full text] [doi: 10.21037/apm.2019.10.01] [Medline: 31735048]
8. León-Salas B, González-Hernández Y, Infante-Ventura D, de Armas-Castellano A, García-García J, García-Hernández M, et al. Telemedicine for neurological diseases: a systematic review and meta-analysis. Eur J Neurol 2023;30(1):241-254. [doi: 10.1111/ene.15599] [Medline: 36256522]
9. Schootemeijer S, van der Kolk NM, Ellis T, Mirelman A, Nieuwboer A, Nieuwhof F, et al. Barriers and motivators to engage in exercise for persons with Parkinson's disease. J Parkinsons Dis 2020;10(4):1293-1299 [FREE Full text] [doi: 10.3233/JPD-202247] [Medline: 32925106]
10. Shalash A, Spindler M, Cubo E. Global perspective on telemedicine for Parkinson's disease. J Parkinsons Dis 2021;11(s1):S11-S18 [FREE Full text] [doi: 10.3233/JPD-202411] [Medline: 33579872]
11. Bashshur R, Shannon G, Krupinski E, Grigsby J. The taxonomy of telemedicine. Telemed J E Health 2011;17(6):484-494 [FREE Full text] [doi: 10.1089/tmj.2011.0103] [Medline: 21718114]
12. Telemedicine: opportunities and developments in member states: report on the second global survey on EHealth. World Health Organization (Global Observatory for eHealth). 2010. URL: https://apps.who.int/iris/handle/ [accessed 2024-12-26]
13. Lindeman D. Position Paper: mHealth technologies: applications to beneft older adults (Discussion Draft). Center for Technology and Aging. Oakland, CA; 2011. URL: https://www.phi.org/thoughtleadership/position-paper-mhealth-technologies-applications-tobenefit-older-adults-discussion-draft/ [accessed 2024-12-26]
14. Samii A, Ryan-Dykes P, Tsukuda RA, Zink C, Franks R, Nichol WP. Telemedicine for delivery of health care in Parkinson's disease. J Telemed Telecare 2006;12(1):16-18. [doi: 10.1258/135763306775321371] [Medline: 16438773]

15. Beck CA, Beran DB, Biglan KM, Boyd CM, Dorsey ER, Schmidt PN, Connect Parkinson Investigators. National randomized controlled trial of virtual house calls for Parkinson disease. Neurology 2017;89(11):1152-1161 [FREE Full text] [doi: 10.1212/WNL.0000000000004357] [Medline: 28814455]
16. Hanson RE, Truesdell M, Stebbins GT, Weathers AL, Goetz CG. Telemedicine vs office visits in a movement disorders clinic: comparative satisfaction of physicians and patients. Mov Disord Clin Pract 2019;6(1):65-69 [FREE Full text] [doi: 10.1002/mdc3.12703] [Medline: 30746418]
17. Schneider RB, Biglan KM. The promise of telemedicine for chronic neurological disorders: the example of Parkinson's disease. Lancet Neurol 2017;16(7):541-551. [doi: 10.1016/S1474-4422(17)30167-9] [Medline: 28566190]
18. Kumar A. Experience of video consultation during the COVID-19 pandemic in elderly population for Parkinson's disease and movement disorders. Postgrad Med J 2021;97(1144):117-118 [FREE Full text] [doi: 10.1136/postgradmedj-2020-138846] [Medline: 33008959]
19. Sibley KG, Girges C, Hoque E, Foltynie T. Video-based analyses of Parkinson's disease severity: a brief review. J Parkinsons Dis 2021;11(s1):S83-S93 [FREE Full text] [doi: 10.3233/JPD-202402] [Medline: 33682727]
20. Wilkinson JR, Spindler M, Wood SM, Marcus SC, Weintraub D, Morley JF, et al. High patient satisfaction with telehealth in Parkinson disease: a randomized controlled study. Neurol Clin Pract 2016;6(3):241-251 [FREE Full text] [doi: 10.1212/CPJ.0000000000000252] [Medline: 27347441]
21. Peacock D, Baumeister P, Monaghan A, Siever J, Yoneda J, Wile D. Perception of healthcare access and utility of telehealth among Parkinson's disease patients. Can J Neurol Sci 2020;47(5):700-704. [doi: 10.1017/cjn.2020.99] [Medline: 32450924]
22. Feeney MP, Xu Y, Surface M, Shah H, Vanegas-Arroyave N, Chan AK, et al. The impact of COVID-19 and social distancing on people with Parkinson's disease: a survey study. NPJ Parkinsons Dis 2021;7(1):10 [FREE Full text] [doi: 10.1038/s41531-020-00153-8] [Medline: 33479241]
23. Siegert C, Hauptmann B, Jochems N, Schrader A, Deck R. ParkProTrain: an individualized, tablet-based physiotherapy training programme aimed at improving quality of life and participation restrictions in PD patients - a study protocol for a quasi-randomized, longitudinal and sequential multi-method study. BMC Neurol 2019;19(1):143 [FREE Full text] [doi: 10.1186/s12883-019-1355-x] [Medline: 31238908]
24. Dobkin RD, Interian A, Durland JL, Gara MA, Menza MA. Personalized telemedicine for depression in Parkinson's disease: a pilot trial. J Geriatr Psychiatry Neurol 2018;31(4):171-176. [doi: 10.1177/0891988718783274] [Medline: 29945467]
25. Jitkritsadakul O, Rajalingam R, Toenjes C, Munhoz RP, Fasano A. Tele-health for patients with deep brain stimulation: the experience of the Ontario Telemedicine Network. Mov Disord 2018;33(3):491-492. [doi: 10.1002/mds.27230] [Medline: 29119600]
26. Chen YY, Guan BS, Li ZY, Yang QH, Xu TJ, Li HB, et al. Application of telehealth intervention in Parkinson's disease: a systematic review and meta-analysis. J Telemed Telecare 2018;26(1-2):3-13. [doi: 10.1177/1357633x18792805]
27. Federico S, Cacciante L, Cieślik B, Turolla A, Agostini M, Kiper P, RIN_TR_Group. Telerehabilitation for neurological motor impairment: a systematic review and meta-analysis on quality of life, satisfaction, and acceptance in stroke, multiple sclerosis, and Parkinson's disease. J Clin Med 2024;13(1) [FREE Full text] [doi: 10.3390/jcm13010299] [Medline: 38202306]
28. Pastana Ramos LF, Vilacorta-Pereira TDCS, Duarte JDS, Yamada ES, Santos-Lobato BL. Feasibility and effectiveness of a remote individual rehabilitation program for people with Parkinson's disease living in the Brazilian Amazon: a randomized clinical trial. Front Neurol 2023;14:1244661 [FREE Full text] [doi: 10.3389/fneur.2023.1244661] [Medline: 37693755]
29. Kraepelien M, Schibbye R, Månsson K, Sundström C, Riggare S, Andersson G, et al. Individually tailored internet-based cognitive-behavioral therapy for daily functioning in patients with Parkinson's disease: a randomized controlled trial. J Parkinsons Dis 2020;10(2):653-664 [FREE Full text] [doi: 10.3233/JPD-191894] [Medline: 32176657]
30. Dobkin RD, Mann SL, Weintraub D, Rodriguez KM, Miller RB, St Hill L, et al. Innovating Parkinson's care: a randomized controlled trial of telemedicine depression treatment. Mov Disord 2021;36(11):2549-2558. [doi: 10.1002/mds.28548] [Medline: 33710659]
31. Maggio MG, Luca A, Cicero CE, Calabrò RS, Drago F, Zappia M, et al. Effectiveness of telerehabilitation plus virtual reality (Tele-RV) in cognitive e social functioning: a randomized clinical study on Parkinson's disease. Parkinsonism Relat Disord 2024;119:105970. [doi: 10.1016/j.parkreldis.2023.105970] [Medline: 38142630]
32. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009;339:b2700 [FREE Full text] [doi: 10.1136/bmj.b2700] [Medline: 19622552]
33. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
34. Cubo E, Mariscal N, Solano B, Becerra V, Armesto D, Calvo S, et al. Prospective study on cost-effectiveness of home-based motor assessment in Parkinson's disease. J Telemed Telecare 2017;23(2):328-338. [doi: 10.1177/1357633X16638971] [Medline: 27000142]
35. Del Pino R, Ortiz de Echevarría A, Díez-Cirarda M, Ustarroz-Aguirre I, Caprino M, Liu J. Virtual coach and telerehabilitation for Parkinson´s disease patients: vCare system. J Public Health 2017;33:1583-1596 [FREE Full text] [doi: 10.1371/journal.pone.0326705] [Medline: 40880394]

36. Dobkin RD, Mann SL, Gara MA, Interian A, Rodriguez KM, Menza M. Telephone-based cognitive behavioral therapy for depression in Parkinson disease: a randomized controlled trial. Neurology 2020;94(16):e1764-e1773 [FREE Full text] [doi: 10.1212/WNL.0000000000009292] [Medline: 32238507]
37. Duffley G, Lutz BJ, Szabo A, Wright A, Hess CW, Ramirez-Zamora A, et al. Home health management of Parkinson disease deep brain stimulation: a randomized clinical trial. JAMA Neurol 2021;78(8):972-981 [FREE Full text] [doi: 10.1001/jamaneurol.2021.1910] [Medline: 34180949]
38. Eldemir S, Guclu-Gunduz A, Eldemir K, Saygili F, Yilmaz R, Akbostancı MC. The effect of task-oriented circuit training-based telerehabilitation on upper extremity motor functions in patients with Parkinson's disease: a randomized controlled trial. Parkinsonism Relat Disord 2023;109:105334. [doi: 10.1016/j.parkreldis.2023.105334] [Medline: 36917914]
39. Ellis TD, Cavanaugh JT, DeAngelis T, Hendron K, Thomas CA, Saint-Hilaire M, et al. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: randomized controlled pilot study. Phys Ther 2019;99(2):203-216. [doi: 10.1093/ptj/pzy131] [Medline: 30715489]
40. Gandolfi M, Geroin C, Dimitrova E, Boldrini P, Waldner A, Bonadiman S, et al. Virtual reality telerehabilitation for postural instability in Parkinson's disease: a multicenter, single-blind, randomized, controlled trial. Biomed Res Int 2017;2017:7962826 [FREE Full text] [doi: 10.1155/2017/7962826] [Medline: 29333454]
41. Goffredo M, Baglio F, DE Icco R, Proietti S, Maggioni G, Turolla A, RIN_TR_Group. Efficacy of non-immersive virtual reality-based telerehabilitation on postural stability in Parkinson's disease: a multicenter randomized controlled trial. Eur J Phys Rehabil Med 2023;59(6):689-696 [FREE Full text] [doi: 10.23736/S1973-9087.23.07954-6] [Medline: 37847247]
42. Heldman DA, Harris DA, Felong T, Andrzejewski KL, Dorsey ER, Giuffrida JP, et al. Telehealth management of Parkinson's disease using wearable sensors: an exploratory study. Digit Biomark 2017;1(1):43-51 [FREE Full text] [doi: 10.1159/000475801] [Medline: 29725667]
43. Theodoros DG, Hill AJ, Russell TG. Clinical and quality of life outcomes of speech treatment for Parkinson's disease delivered to the home via telerehabilitation: a noninferiority randomized controlled trial. Am J Speech Lang Pathol 2016;25(2):214-232. [doi: 10.1044/2015_AJSLP-15-0005] [Medline: 27145396]
44. Esper CD, Valdovinos BY, Schneider RB. The importance of digital health literacy in an evolving Parkinson's disease care system. J Parkinsons Dis 2024;14(s1):S181-S189 [FREE Full text] [doi: 10.3233/JPD-230229] [Medline: 38250786]
45. Ruotolo I, Sellitto G, Berardi A, Simeon R, Panuccio F, Amadio E, et al. Psychometric properties of the Parkinson's Disease Questionnaire-39 and its short form Parkinson's disease Questionnaire-8: a systematic review and meta-analysis. J Clin Neurosci 2024;123:100-117 [FREE Full text] [doi: 10.1016/j.jocn.2024.03.032] [Medline: 38564966]
46. Tuba BK, Elif DE, Ozgur O, Fatos ES, Emine KAE, Ozden TO. Reliability and validity of the Turkish version of the 39-item Parkinson Disease Questionnaire. Ideggyogy Sz 2023;76(5-6):181-188 [FREE Full text] [doi: 10.18071/isz.76.0181] [Medline: 37294025]
47. Diwakar V, Ertmer PA, Nour AYM. Developing interactive course Web sites for distance education and characteristics of students enrolled in distance learning courses. J Vet Med Educ 2003;30(4):351-357. [doi: 10.3138/jvme.30.4.351] [Medline: 14976621]
48. Li J, Liu Y, Jiang J, Peng X, Hu X. Effect of telehealth interventions on quality of life in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Int J Nurs Stud 2021;122:103970. [doi: 10.1016/j.ijnurstu.2021.103970] [Medline: 34303269]
49. Xu Y, Feeney MP, Surface M, Novak D, Troche MS, Beck JC, et al. Attitudes toward telehealth services among people living with Parkinson's disease: a survey study. Mov Disord 2022;37(6):1289-1294 [FREE Full text] [doi: 10.1002/mds.28990] [Medline: 35338664]
50. Chaudhuri KR, Azulay J, Odin P, Lindvall S, Domingos J, Alobaidi A, et al. Economic burden of Parkinson's disease: a multinational, real-world, cost-of-illness study. Drugs Real World Outcomes 2024;11(1):1-11 [FREE Full text] [doi: 10.1007/s40801-023-00410-1] [Medline: 38193999]
51. Brock P, Oates LL, Gray WK, Henderson EJ, Mann H, Haunton VJ, et al. Driving and Parkinson's disease: a survey of the patient's perspective. J Parkinsons Dis 2022;12(1):465-471 [FREE Full text] [doi: 10.3233/JPD-212686] [Medline: 34542030]
52. Zhang P, Jin W, Lyu Z, Lyu X, Li L. Study on the mechanism of gut microbiota in the pathogenetic interaction between depression and Parkinson 's disease. Brain Res Bull 2024;215:111001 [FREE Full text] [doi: 10.1016/j.brainresbull.2024.111001] [Medline: 38852651]
53. Tolin DF, Lord KA, Knowles KA. Cognitive-behavioral therapy enhancement strategies. Psychiatr Clin North Am 2024;47(2):355-365. [doi: 10.1016/j.psc.2024.02.005] [Medline: 38724125]
54. Xiang X, Kayser J, Turner S, Ash S, Himle JA. Layperson-supported, web-delivered cognitive behavioral therapy for depression in older adults: randomized controlled trial. J Med Internet Res 2024;26:e53001 [FREE Full text] [doi: 10.2196/53001] [Medline: 38437013]
55. Dou J, Wang J, Gao X, Wang G, Bai Y, Liang Y, et al. Effectiveness of telemedicine interventions on motor and nonmotor outcomes in Parkinson disease: systematic review and network meta-analysis. J Med Internet Res 2025;27:e71169 [FREE Full text] [doi: 10.2196/71169] [Medline: 40460428]

56. Santos DT, Camelo DMF, Strelow MZ, Silva MTS, Führ P, Marins LW, et al. Feasibility of telemedicine for patients with parkinsonism in the Brazilian public health system. Arquivos de neuro-psiquiatria 2022;80(9):914-921 [FREE Full text] [doi: 10.1055/s-0042-1755323] [Medline: 36261128]
57. Lima DP, Gomes VC, Viana Júnior AB, Assis F, Oliveira P, Cunha L, Chaves, et al. Telehealth for Parkinson disease patients during the COVID-19 pandemic: the TeleParkinson study. Arq Neuropsiquiatr 2022;80(10):1026-1035 [FREE Full text] [doi: 10.1055/s-0042-1758751] [Medline: 36535287]
58. Joo JY, Yun JY, Kim YE, Jung YJ, Kim R, Yang H, et al. A survey of perspectives on telemedicine for patients with Parkinson's disease. J Mov Disord 2024;17(1):89-93 [FREE Full text] [doi: 10.14802/jmd.23130] [Medline: 37604653]
59. Perskaudas R, Myers CE, Interian A, Gluck MA, Herzallah MM, Baum A, et al. Reward and punishment learning as predictors of cognitive behavioral therapy response in Parkinson's disease comorbid with clinical depression. J Geriatr Psychiatry Neurol 2024;37(4):282-293. [doi: 10.1177/08919887231218753] [Medline: 38158704]
60. Rohrmann T, Praus P, Proctor T, Benedyk A, Tost H, Hennig O, et al. Patients with affective disorders profit most from telemedical treatment: evidence from a naturalistic patient cohort during the COVID-19 pandemic. Front Psychiatry 2022;13 [FREE Full text] [doi: 10.3389/fpsyt.2022.971896] [Medline: 36532188]
61. Angelopoulou E, Papageorgiou SG. Telemedicine in Alzheimer's disease and other dementias: where we are? J Alzheimers Dis 2025;103(1):3-18. [doi: 10.1177/13872877241298295] [Medline: 39639574]
62. Lau TK, Tse M, Liu Y, Leung AYM. Effectiveness of technological interventions on psychosocial well-being and perception of technological interventions among people with Parkinson's disease: a systematic review. Australas J Ageing 2025;44(2):e70034. [doi: 10.1111/ajag.70034] [Medline: 40317851]
63. Lee D, Erande A, Christodoulou G, Malik S. Addressing mental health symptoms among COVID-19 healthcare workers: a heart rate variability biofeedback pilot study. Stress Health 2024;40(6):e3502 [FREE Full text] [doi: 10.1002/smi.3502] [Medline: 39513426]
64. García-Bustillo Á, Ramírez-Sanz JM, Garrido-Labrador JL, Olivares-Gil A, Valiñas-Sieiro F, Allende-Río M, et al. A multidisciplinary telemedicine approach for managing frailty in Parkinson's disease. A longitudinal, case-control study. Parkinsonism Relat Disord 2024;130:107215 [FREE Full text] [doi: 10.1016/j.parkreldis.2024.107215] [Medline: 39586130]
65. Roper A, Brooks D, Mitchell LK, Pachana NA, Au TR, Byrne GJ, et al. Feasibility and acceptability of a videoconferencing CBT intervention for anxiety in people with Parkinson's disease. Clin Gerontol 2025;48(4):828-843 [FREE Full text] [doi: 10.1080/07317115.2024.2306861] [Medline: 38277135]
66. Wuthrich VM, Rapee RM. Telephone-delivered cognitive behavioural therapy for treating symptoms of anxiety and depression in Parkinson's disease: a pilot trial. Clin Gerontol 2019;42(4):444-453. [doi: 10.1080/07317115.2019.1580811] [Medline: 30821649]
67. Poon S, Tan C, Hong W, Chen KC, Yu R. Tailoring anxiety assessment for Parkinson's disease: the Chinese Parkinson anxiety scale with cultural and situational anxiety considerations. Soc Sci Med 2025;381 [FREE Full text] [doi: 10.1016/j.socscimed.2025.118284] [Medline: 40513505]
68. What is Parkinson?s? 2017. Association EPsD. URL: http://www.epda.eu.com/about-parkinsons/what-is-parkinsons/ [accessed 2026-01-06]
69. Barbour PJ, Arroyo J, High S, Fichera LB, Staska-Pier MM, McMahon MK. Telehealth for patients with Parkinson's disease: delivering efficient and sustainable long-term care. Hosp Pract (1995) 2016;44(2):92-97. [doi: 10.1080/21548331.2016.1166922] [Medline: 26982525]
70. Ekmekyapar Fırat Y, Turgay T, Soğan SS, Günel Karadeniz P. Effects of LSVT-BIG via telerehabilitation on non-motor and motor symptoms and quality of life in Parkinson's disease. Acta Neurol Belg 2023;123(1):207-214 [FREE Full text] [doi: 10.1007/s13760-022-02104-x] [Medline: 36175786]
71. Mustile M, Kourtis D, Ladouce S, Edwards MG, Volpe D, Pilleri M, et al. Investigating the brain mechanisms of externally cued sit-to-stand movement in Parkinson's disease. Mov Disord 2024;39(9):1556-1566. [doi: 10.1002/mds.29889] [Medline: 38984716]
72. D'Souza AF, Jasti DB, Rao RR, Natarajan M. Feasibility of a tele-assisted home exercise program for balance and functional mobility in persons with Parkinson's disease (TELEPORT-PD). Int J Telemed Appl 2025;2025:9936329 [FREE Full text] [doi: 10.1155/ijta/9936329] [Medline: 40519980]
73. Kumar A, Patil S, Singh VK, Pathak A, Chaurasia RN, Mishra VN, et al. Assessment of non-motor symptoms of Parkinson's disease and their impact on the quality of life: an observational study. Ann Indian Acad Neurol 2022;25(5):909-915 [FREE Full text] [doi: 10.4103/aian.aian_647_21] [Medline: 36561034]
74. Diaconu, Irincu L, Ungureanu L, Ciopleia B, în D, Falup-Pecurariu C. Restless legs syndrome in Parkinson's disease. J Pers Med 2023;13(6) [FREE Full text] [doi: 10.3390/jpm13060915] [Medline: 37373904]
75. Ha J, Park JH, Lee JS, Kim HY, Song JO, Yoo J, et al. Effectiveness of live-streaming tele-exercise intervention in patients with Parkinson's disease: a pilot study. J Mov Disord 2024;17(2):189-197 [FREE Full text] [doi: 10.14802/jmd.23251] [Medline: 38419488]
76. Fleisher JE, Hess SP, Klostermann EC, Lee J, Myrick E, Mitchem D, et al. IN-HOME-PD: the effects of longitudinal telehealth-enhanced interdisciplinary home visits on care and quality of life for homebound individuals with Parkinson's

disease. Parkinsonism Relat Disord 2022;102:68-76 [FREE Full text] [doi: 10.1016/j.parkreldis.2022.07.017] [Medline: 35963046]
77. Bode M, Kalbe E, Liepelt-Scarfone I. Cognition and activity of daily living function in people with Parkinson's disease. J Neural Transm (Vienna) 2024;131(10):1159-1186. [doi: 10.1007/s00702-024-02796-w] [Medline: 38976044]
78. Latella D, Maresca G, Formica C, Sorbera C, Bringandì A, Di Lorenzo G, et al. The role of telemedicine in the treatment of cognitive and psychological disorders in Parkinson's disease: an overview. Brain Sci 2023;13(3) [FREE Full text] [doi: 10.3390/brainsci13030499] [Medline: 36979309]
79. Johnson JK, Longhurst JK, Gevertzman M, Jefferson C, Linder SM, Bethoux F, et al. The use of telerehabilitation to improve movement-related outcomes and quality of life for individuals with Parkinson disease: pilot randomized controlled trial. JMIR Form Res 2024;8:e54599 [FREE Full text] [doi: 10.2196/54599] [Medline: 39083792]
80. Putzolu M, Manzini V, Gambaro M, Cosentino C, Bonassi G, Botta A, et al. Home-based exercise training by using a smartphone app in patients with Parkinson's disease: a feasibility study. Front Neurol 2023;14:1205386 [FREE Full text] [doi: 10.3389/fneur.2023.1205386] [Medline: 37448748]
81. Cubo E, Delgado-López PD. Telemedicine in the management of Parkinson's disease: achievements, challenges, and future perspectives. Brain Sci 2022;12(12) [FREE Full text] [doi: 10.3390/brainsci12121735] [Medline: 36552194]
82. Kotsimpou S, Liampas I, Dastamani M, Marogianni C, Stamati P, Tsika A, et al. Evaluation of computer-based cognitive training on mild cognitive impairment in Parkinson's disease (PD-MCI): a review. J Clin Med 2025;14(9) [FREE Full text] [doi: 10.3390/jcm14093001] [Medline: 40364031]
83. Tagliente S, Minafra B, Aresta S, Santacesaria P, Buccoliero A, Palmirotta C, et al. Effectiveness of a home-based computerized cognitive training in Parkinson's disease: a pilot randomized cross-over study. Front Psychol 2024;15 [FREE Full text] [doi: 10.3389/fpsyg.2024.1531688] [Medline: 39850970]
84. Isernia S, Di Tella S, Rossetto F, Borgnis F, Realdon O, Cabinio M, et al. Exploring cognitive reserve's influence: unveiling the dynamics of digital telerehabilitation in Parkinson's disease resilience. NPJ Digit Med 2024;7(1):116 [FREE Full text] [doi: 10.1038/s41746-024-01113-9] [Medline: 38710915]
85. D'Iorio A, Aiello E, Vitale C, Amboni M, Verde F, Silani V, et al. Diagnostics and ecological validity of the Italian version of the Parkinson's disease cognitive rating scale. Dement Geriatr Cogn Disord 2025;54(5):347-351. [doi: 10.1159/000545090] [Medline: 40068648]

Abbreviations

ADL: activities of daily living
BBQ: Brunnsviken Brief Quality of Life Scale
BDI: Beck Depression Inventory
CBT: cognitive behavioral therapy
HADS-A: Hospital Anxiety and Depression Scale-Anxiety
HADS-D: Hospital Anxiety and Depression Scale-Depression
HAM-A: Hamilton Anxiety Rating Scale
HAM-D: Hamilton Depression Rating Scale
HKSJ: Hartung-Knapp-Sidik-Jonkman
MDS-UPDRS: Movement Disorders Society-Unified Parkinson’s Disease Rating Scale
MMSE: Mini-Mental State Examination
MoCA: Montreal Cognitive Assessment
PD: Parkinson disease
PDQ: Parkinson Disease Questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QOL: quality of life
RCT: randomized controlled trial
SF-36: Medical Outcomes Study 36-Item Short Form Health Survey
SMD: standard mean difference

Edited by A Stone; submitted 07.Jan.2025; peer-reviewed by F Asci, S Tonapa, K Elechi; comments to author 05.May.2025; revised version received 29.Dec.2025; accepted 30.Dec.2025; published 28.Jan.2026.

Please cite as:
Sun M, Tang F, min L, Wen S, Wang S, Jiang H
Effects of Telehealth Interventions for People With Parkinson Disease: Systematic Review and Meta-Analysis of Randomized Controlled Trials
JMIR Mhealth Uhealth 2026;14:e70994
URL: https://mhealth.jmir.org/2026/1/e70994
doi:10.2196/70994
PMID:

Original Paper

The Landscape of Mobile Apps for Healthy Eating: Case Study for a Systematic Review and Quality Assessment

Garlene Zamora Zamorano[1*], RDN; Alejandro Déniz-García[1,2*], MD; Alezandra Torres-Castaño[3], PhD; María Luisa Álvarez-Malé[1], PhD; Inger Torhild Gram[4,5], MD, PhD; Guri Skeie[5,6], PhD; Ana M Wägner[1,2], MD, PhD

[1]Instituto de Investigaciones Biomédicas y Sanitarias, Universidad de Las Palmas de Gran Canaria, Las Palmas de Gran Canaria, Spain

[2]Department of Endocrinology and Nutrition, Complejo Hospitalario Universitario Insular Materno-Infantil de Canarias, Instituto de Investigaciones Sanitarias de Canarias, Las Palmas de Gran Canaria, Spain

[3]Evaluation Unit of the Canary Islands Health Service, Instituto de Investigación Sanitaria de Canarias, Tenerife, Spain

[4]Norwegian Centre for eHealth Research, University Hospital of North Norway, Tromsø, Norway

[5]Department of Community Medicine, Faculty of Health Sciences, UiT the Arctic University of Norway, Tromsø, Norway

[6]Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France

[*]these authors contributed equally

Corresponding Author:
Ana M Wägner, MD, PhD
Instituto de Investigaciones Biomédicas y Sanitarias
Universidad de Las Palmas de Gran Canaria
P Blas Cabrera Felipe “Físico” (s/n)
Las Palmas de Gran Canaria, 35016
Spain
Phone: 34 928453431
Email: ana.wagner@ulpgc.es

Abstract

Background: Mobile apps are being increasingly used to foster healthy lifestyles. There is a growing need for clear, standardized guidelines to help users select safe and effective health apps.

Objective: Our study aimed to highlight the importance of establishing a structured framework for quality evaluation in mobile health (mHealth) through a case study of mobile apps promoting healthy eating.

Methods: We conducted a systematic review of apps promoting healthy eating that had already been evaluated by one or more of 28 recognized health app certification bodies. Three rounds of app evaluations were conducted by experts in nutrition and behavior change. The first two rounds focused on the quality of the content of the recommendations and were performed pairwise using the Quality Evaluation Scoring Tool (QUEST), which has not been previously used by the certification bodies. In addition, in the second and third rounds, each reviewer answered the question “How probable is it that you would recommend this app?” using a subjective scale score from 0 to 10. In the third round, this score was weighed by usability (30%), content quality (40%), and promotion of behavior change (30%). Discussions were held to resolve scoring discrepancies and to identify the top-quality apps. We also assessed correlations among QUEST, Google Play Store, and certification body scores.

Results: Of the 41 apps identified by five certification bodies, 19 (46.3%) met the inclusion criteria and were examined. Only 16 (84.2%) of these remained accessible for the second round. Eight of these surpassed 20 points (out of a maximum of 28) on the QUEST scale and were evaluated by all six experts in the third round, and the top 5 (62.5%) apps were selected. No correlations were found among QUEST, Google Play Store, and certification body scores.

Conclusions: Despite numerous evaluations by various certification bodies, only 5 (12.2%) of the 41 apps met the quality standards set by our experts. Our results mark the importance of rigorous, transparent, and standardized app evaluation processes to guide users toward making informed decisions about health apps. Guidelines for developers for the design of evidence-based, unbiased, high-quality apps, as well as technological solutions for real-time monitoring of the health apps, would address these challenges and improve reliability.

(JMIR Mhealth Uhealth 2026;14:e68737) doi:10.2196/68737

KEYWORDS

apps; applications; mobile health; mHealth; nutrition; healthy eating; eHealth; digital health

Introduction

As our world becomes increasingly digital, the use of mobile apps for health-related purposes (health apps) is on the rise [1,2]. The World Health Organization's (WHO) Global Observatory of eHealth describes health apps as key components of the broader domain of mobile health (mHealth). The use of mobile devices (eg, smartphones, patient-monitoring tools, personal digital assistants, and other wireless devices) is instrumental to support medical and public health practices. eHealth encompasses the cost-effective and secure use of information and communication technologies for health care and related fields [3]. Since WHO's acknowledgment in 2016 [4], the market has expanded rapidly. There has been an exponential increase in health app availability from 325,000 in 2017 [5] to an estimated 255 billion app downloads in 2022 [1,2].

Despite this growth, health care professionals lack unified standards for evaluating health app quality, safety, and effectiveness [6]. Recent European Union regulations on medical devices, such as Regulation (EU) 2017/745 enacted in May 2021, represent significant development. These regulations classify certain health apps as medical devices, necessitating adherence to specific criteria for approval. Supplementary documents from the European Commission [7,8] complement these regulations. The regulations also establish a medical device database to enhance transparency for both patients and health care providers. However, despite mHealth apps being within the medical device framework and subject to all corresponding laws at the European level, there are pending challenges related to user data management. Commercial platforms typically do not provide reliable information regarding the efficacy or safety of the apps they offer. Health apps should be substantiated by scientific evidence, facilitating their endorsement by health care professionals and enabling end users to benefit from a validated certification system [9,10], which should include the assessment of not only the quality of the information provided but also other relevant factors associated with app quality, such as usability and behavior change promotion [11].

This case study proposed a robust process for the evaluation of health apps, using dietary intervention apps promoting healthy eating as an example, to illustrate both the evaluation method and the challenges faced by end users and health care prescribers. A particular focus was placed on the quality of the health information provided by the apps.

Methods

Study Design

Our approach was structured in two stages: first, the systematic identification of relevant apps and, second, their assessment based on predefined quality, usability, and evidence-based criteria. This sequence aims to be a realistic, reproducible way to support both health care professionals and end users in the navigation and evaluation of the growing mHealth landscape.

App Identification

An initial exploratory analysis was conducted to identify projects, initiatives, and organizations involved in the evaluation of health apps. Most of these resources have been catalogued in two documents by the Spanish Ministry of Health [12] and the European project mHealth-Hub [13,14]. Both include governmental and nongovernmental efforts. Each initiative was then individually reviewed to examine its content and structure. Notable heterogeneity among the initiatives was observed, allowing them to be classified into two main groups: (1) platforms offering pre-evaluated apps, used as databases for the search, and (2) those focused on normative aspects, standardization, and quality assessment. Given the diversity in the approaches and structures of these platforms, the search began with general terms related to the study's focus, such as healthy eating, staying healthy, diet, lifestyle, preventive medicine, weight or healthy habits, which were later adapted to the navigation logic, taxonomy, or filtering system of each initiative (for more details, see Multimedia Appendix 1).

From August to December 2021, we searched for health apps that met the following inclusion criteria: aimed at adults (over 18 years old), available in English or Spanish, and offering nutritional guidance (eg, recommendations for dietary changes or meal planning). Apps that solely functioned as food diaries, calorie counters, or barcode readers but did not include any specific recommendations were excluded. We examined app descriptions and features to confirm eligibility.

In the absence of established guidelines for health app evaluation, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Multimedia Appendix 2) [15,16] adapted to our research objectives. We reviewed 28 certification bodies (see Table S1 in Multimedia Appendix 1), of which only 5 (17.9%) contained apps meeting our inclusion criteria (see Table 1 for additional details):

Table 1. Summary of certification/assessment bodies that included apps promoting healthy eating.

Name	Year it started	Country	Reviewers	Evaluation criteria	Public or private
MyHealthApps[a]	2013	United Kingdom	End users, associations, or caretakers	Scoring system	Private
Healthy Living Apps Guide	2018	Australia	Expert reviewers (at least two on behavior change/public health)	MARS[b] (functionality) and ABACUS[c] (behavior change)	Public (Government of Victoria)
ORCHA[d]	2018	United Kingdom	Experts and reviewers (end users considered)	A 7-step system	Private (cooperation with the NHS[e])
GGD[f] AppStore	2016	Netherlands	Experts and end users	A 4-step assessment based on questionnaires, including a final evaluation of behavior change techniques used	Public (GGD GHOR[g])
Health Navigator	2017	New Zealand	Experts, clinical reviewers, and end users	Internal revision by experts, clinical revision, and end-user revision.	Public Ministry of Health

[a]Closed on May 17, 2022.

[b]MARS: Mobile App Rating Scale.

[c]ABACUS: App Behavior Change Scale.

[d]ORCHA: Organization for the Review of Care and Health Applications.

[e]NHS: National Health Service.

[f]GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service).

[g]GHOR: Geneeskundige Hulpverleningsorganisatie in de Regio (Regional Medical Emergency Preparedness and Planning).

Healthy Living AppsGuide [17] assesses app functionality using the Mobile App Rating Scale (MARS) [18] and behavior change using the App Behavior Change Scale (ABACUS) [19]. Each criterion is rated on a scale of 0-5 stars, with additional considerations, such as app cost and data export capabilities.

MyHealthApps [20] focused on user preferences and developer-related data. This resource operated as a catalog, rather than assigning individual scores to apps. It assessed transparency regarding pricing, contact details, geographical location, and security measures. Unfortunately, it closed in May 2022.

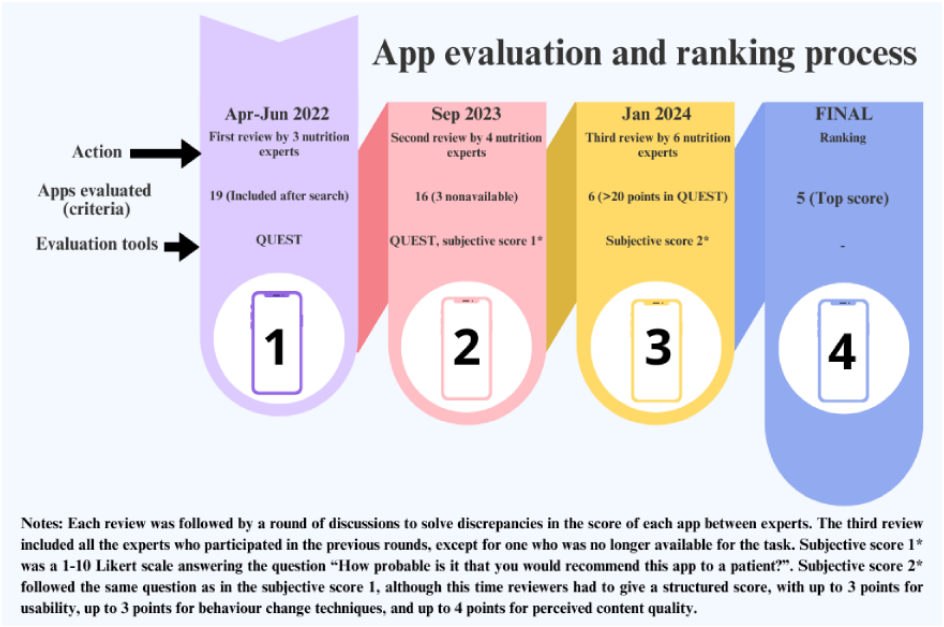
ORCHA (Organization for the Review of Care and Health Applications) [21] uses a seven-step evaluation system comprising three main domains: data, professional guarantee, and usability and accessibility. App functionalities and features are also considered. A maximum score of 100 points is attainable for each domain, with ratings above 65 considered good quality. Scores between 45 and 65 suggest areas for further investigation, and scores below 45 deem the app (or domain) potentially unsafe or ineffective.

Gemeentelijke Gezondheidsdienst (*GGD*) *AppStore* [22] follows an evaluation method considering app availability, pricing, compatibility with different operating systems, provider contact details, promotion of healthy behavior, data management, and goal setting. To enter evaluation, apps must focus on self-care and include at least two methods for behavior change. The apps are also assessed for usability, reliability, privacy, safety, and relevance, with each characteristic rated as good, sufficient, or inadequate. Apps are scored up to a maximum of 5 stars.

Health Navigator [23] is supported by the New Zealand Ministry of Health and provides a library of reliable apps. The evaluation process encompasses app features, functionalities, quality of information, and relevance to users. Apps must fulfil specific criteria, including evidence-based content and an evaluation of effectiveness, acceptability, and usability, and include a privacy statement. Clinicians with diverse expertise rate the apps from 1 (very poor, not recommended) to a maximum of 5 stars (excellent). End users also provide feedback. Finally, apps deemed clinically unsafe or potentially harmful are excluded.

As health apps are frequently updated, a follow-up search was performed in May 2022. By that time, *MyHealthApps* was no longer active. From this point on, three rounds of app evaluations followed (see Figure 1).

Figure 1. Diagram depicting the evaluation process of the apps.



App Assessment

The apps underwent three rounds of evaluation involving a total of seven reviewers who downloaded and assessed them on the devices described later. An endocrinologist (specialized in endocrinology and nutrition), one dietician, two psychologists, and three epidemiology and public health professionals participated. A second endocrinologist led the evaluation process and coordinated and summarized meetings.

Evaluation Round 1 (April-July 2022)

The content of each of the selected apps was evaluated by pairs from a panel of three nutrition experts (author AD-G, specialized in endocrinology and nutrition; author GZZ, a registered dietician; and Cristina Ruano Rodríguez, a lecturer in nutrition and public health). To facilitate comparison across the apps, we used the Quality Evaluation Scoring Tool (QUEST) [24], not used by any of the evaluating bodies that included the selected apps. QUEST was originally designed to assess the quality of health information in digital media and has been tested for reliability and validity. It offers a total score ranging from 0 to 28 points based on the following criteria: Authorship (0, 1, or 2 points) evaluates the ease of identifying the authors of the content; attribution (0, 3, 6, or 9 points) assesses whether health information is backed by scientific studies; type of study (0, 1, or 2 points, if at least 6 points have been given to *attribution*) evaluates the studies' quality; conflict of interest (0, 3, or 6 points) evaluates whether the information promotes purchase of products or services; complementarity (0 or 1 point) evaluates whether the information supports the health provider-patient relationship; and tone (0, 3, or 6 points) assesses language used as biased, neutral, or acknowledging the limits of knowledge.

The apps were downloaded by the reviewers to Android devices (two Galaxy Tab A7 Lite SM-T220 running on One UI Core 3.1 and Android 11 and a Redmi 9A M2006C3LG MIUI Global 12.0.20 mobile phone running on MIUI Global 12.0.20 Estable [QCDEUXM] and Android 10QP1A.190711.020). Whenever a premium version was available, it also was tested. A data extraction template was developed, including reviewer ID, date of the evaluation, name of the app, developer, version, device where it was downloaded, score for each item in QUEST, and comments. The mean total score was calculated to rank the apps, and inter-reviewer agreement for each QUEST item was measured using weighted κ coefficients. Concordance for the total score was evaluated using the intraclass correlation coefficient. We also explored potential correlations between the QUEST scores, Google Play Store ratings, and the scores from each assessment body (all three, nonnormally distributed quantitative continuous variables) using Spearman's coefficients. For these procedures, we used the *vcd* library in R software (RStudio version 1.3.1056, R Foundation for Statistical Computing). Following the analysis, discrepancies in the evaluation of QUEST items were discussed to identify their underlying causes.

Evaluation Round 2 (September 2023)

An expanded panel of four additional reviewers (authors AT-C and MLA-M, psychologists, experts in behavior change; author ITG, professor in epidemiology, expert in health promotion; and author GS, public health nutritionist, expert in prevention through diet) re-evaluated the apps (4-5 each) using the same methodology but limiting the time spent in the evaluation to a maximum of 45 minutes per app. Apps with a score average over 20 points (out of a maximum of 28) either in the initial evaluation or in the update were selected for further analysis. Although no given cutoff is recommended for the tool, this threshold was chosen to ensure the inclusion of apps that meet a fair-to-high standard. All evaluators were asked to subjectively score (0-10 points) their assigned apps by answering the question "How probable is it that you would recommend this app?" One of the reviewers (AD-G) scored all of them. Apps with a difference of 3 or more points in the initial and updated QUEST scores were discussed within the same reviewer pair (one of the reviewers involved in the first round and the new expert involved in the second round) to solve or explain this discrepancy. This process was documented, summarized, and

then discussed during a meeting involving all the authors, including author AMW (specialist in endocrinology, nutrition, and diabetes). The meeting was recorded, transcribed, and summarized in a document, which was shared with the participants for feedback.

Evaluation Round 3 (January 2024)

The last version of the selected apps was downloaded and assessed by six experts (GZZ, AD-G, AT-C, MLA-M, ITG, and GS), who had participated in the previous rounds. A subjective score was given again, this time assigning up to 4 points for content quality (this aspect being the most relevant for our research), 3 for usability, and 3 for promoting behavior change. The specific question to answer in this round was "How probable is it that you would recommend this app for someone to improve their healthy diet, based on the quality of the content (40%), usability (30%), and promotion of behavior change (30%)?"

Further discussion among all the authors led to the final ranking and the selection of the top 5 apps.

Ethical Considerations

Given the nature of this study, no ethics committee approval or informed consent to participate was needed. All authors declared their consent for publication.

Results

App Details

Of the 28 certification bodies identified (see Table S1 in Multimedia Appendix 1), 14 (50%) catalogued a total of 557 apps across various categories, encompassing terms such as "healthy eating," "nutrition," "diet", "healthy practices," and "lifestyle". Only 5 (17.9%) included apps that fulfilled our inclusion criteria (see Table 1). These certification bodies used a variety of criteria in their evaluation processes, in terms of both the elements assessed and the tools used. Nevertheless, there was a significant overlap in the factors considered key for determining app quality, including functionality, usability, engagement, aesthetics, privacy, data protection, and effectiveness in promoting behavior change.

A total of 41 (7.4%) apps met the inclusion criteria. After eliminating duplicates, 30 (73.2%) apps remained, of which 11 (36.7%) met the exclusion criteria (see Figure 2).

The characteristics of the selected 19 (63.3%) apps are summarized in Table 2. Of these 19 apps, 5 (26.3%) were included in two or more of the certification bodies: *MyFitnessPal*, *FatSecret*, *MyNetDiary*, *Noom*, and *8fit Workout & Meal Planner*. Only the apps included in ORCHA had undergone formal evaluation of their content.

As per the definition, all apps included nutritional interventions, but most also combined different strategies or features, making their categorization challenging. Overall, 14 (73.7%) apps included weight control or tracking (numbered 1-6, 8, 11, 12, 14, 16-19 in Table 2), 12 (63.2%) allowed for calorie counting (numbered 1-4, 6-8, 11-13, 17, 19 in Table 2), and 13 (68.4%) facilitated the recording of physical activity (numbered 1-4, 6-9, 14-17, 19 in Table 2). Furthermore, 12 (63.2%) apps (numbered 1-8, 10, 11, 16, 18 in Table 2) incorporated behavior change techniques, such as motivational feedback, health education promotion, and goal setting.

Figure 2. PRISMA 2020 flowchart describing app search and findings. GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service); ORCHA: Organization for the Review of Care and Health Applications; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

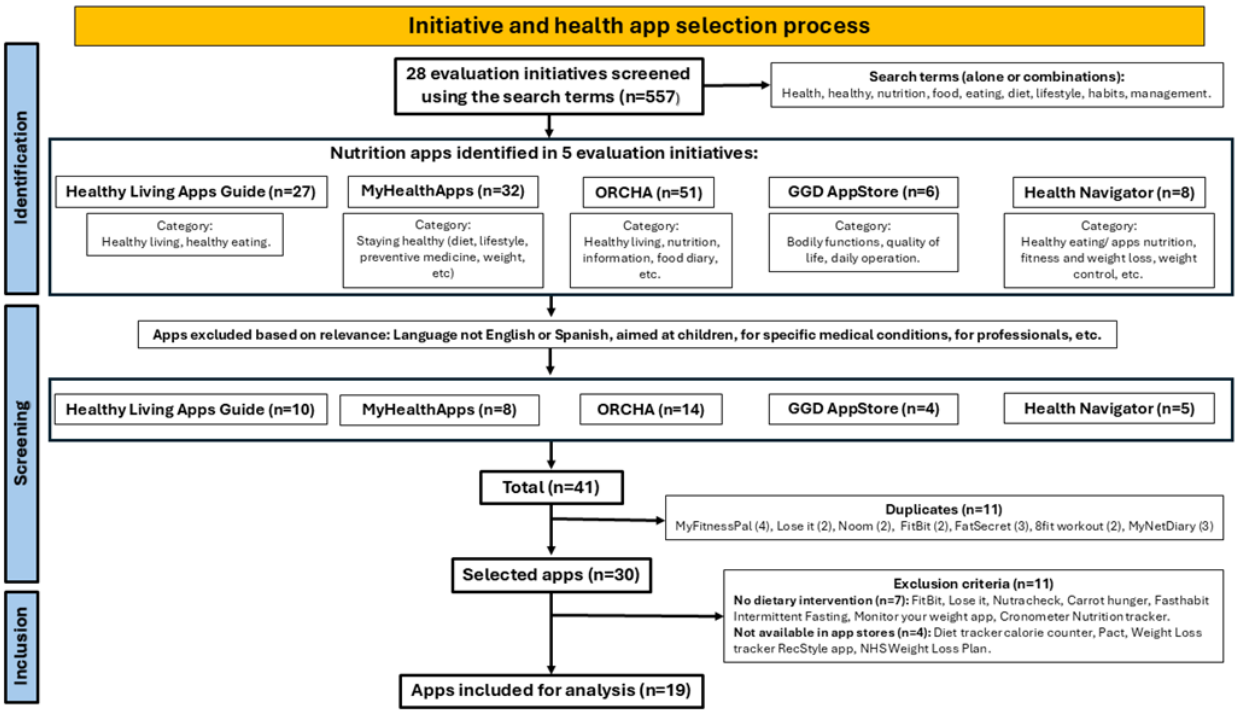


Table 2. Main features of the 19 selected and evaluated health apps.

App number	App name; developer	Description
1	Noom; Noom Inc	It uses behavior change psychology to promote healthier habits, weight loss, and health goals. It can track meals and link to a pedometer.
2	YAZIO; YAZIO	It is a food diary and calorie counter that also offers diets and recipes, including vegan, vegetarian, and intermittent fasting options.
3	FatSecret; FatSecret	It is a food diary and calorie counter. It allows the user to scan products barcodes, tracking weight and keeping a record of meals and consumed food photos.
4	MyNetDiary; MyNetDiary Inc	It is a food log and calorie counter that also provides diets and recipes. It has a food database and a barcode scanner. It also provides motivational messages to achieve nutritional goals.
5	DietBet; WayBetter Inc	It aims to help the user achieve healthy weight loss and maintain motivation. It offers a support community, customer service, and expert-led games (coaches, nutritionists, etc).
6	MyFitnessPal; MyFitnessPal Inc	It offers recipes, a meal planner, a calorie counter, and workout plans. It also allows tracking progress and setting macronutrient goals.
7	MyPlate[a]; LIVESTRONG	It allows tracking calories and exercises. It also tracks the user's progress based on calorie and nutrient goals and provides motivation through support groups.
8	LIFESUM; Lifesum	It functions as a food diary and offers various types of diets and recipes. It also allows tracking exercise and water intake and setting weight goals and health data.
9	8fit Workout & Meal Planner; Urbanite Inc	It offers workout routines in different categories (boxing, yoga, high-intensity interval training, etc), along with meal plans.
10	Freeletics Nutrition; Freeletics GmbH	It acts as a nutrition coach to help the user achieve dietary goals. It guides the user toward healthy eating, recipes, etc, and helps them adapt nutrition to personal goals.
11	GetFit-Daily Meal Planner[a] App Prodakshn;OOO	It provides daily and weekly meal plans to lose, maintain, or gain weight based on the user's goals. It also calculates nutrients, calories, water intake, BMI, etc, and includes reminders.
12	HealthifyMe; HealthifyMe	It can act as a calorie counter, provides weight loss and exercise plans, guides sleep hygiene, etc.
13	Eat This Much-Meal Planner; Eat This Much Inc	It is a meal planner based on dietary preferences, budget, etc, and can be configured for different types of diets.
14	LIFE Extend[a]; LifeOmic	It is marketed as a precision health mobile app to help improve users' health using five pillars of health.
15	Fastic Fasting Ap; Fastic GmbH	It promotes intermittent fasting with different levels of intensity (gentle, moderate, intense).
16	Second Nature; Second Nature	It provides a lifestyle change program that helps lose weight and develop healthy habits. Nutritionists guide the user throughout the program. The app also includes a support group.
17	Uplyfe-Precision Nutrition[a]; Uplyfe AG	It is marketed as a customized nutrition and lifestyle change app-guided program based on scientific findings. It also provides personalized nutrition plans, activities, and symptom tracking.
18	Freshwell; Freshwell	Health care personnel promote a low-carb diet. The app provides informative weekly modules, a meal planner, etc.
19	Contador de calorías; Virtuagym	It provides a personalized nutrition plan according to the user's lifestyle and goals.

[a]No longer available.

App Evaluation

Table 3 presents the scores given to each app by the reviewers during the three rounds of assessment, by the certification bodies, and by a commercial app store.

Table 3. Scores by the different evaluation tools and rounds of revision for each app.

Evaluation rounds and app (name; first version evaluated; date of update of that version; year launched)	Scores								
	Round 1: QUEST[a]	Round 2: QUEST	Round 2: subjective	Round 3: subjective	Google Play Store (stars); August 19, 2022	OR-CHA[b]	Health Navigator	GGD[c] App-Store	Healthy Living Apps Guide
Maximum score	28	28	10	10	5	100	5	5	5
Number of evaluators (total, per app)	3, 2	4, 1	6, 6	6, 6	__[d]	—	—	—	—
Apps that went through all evaluation rounds									
Freshwell; 1.1.1; October 25, 2021; 2021	21.5	28	7.2	6.9	5	68	—	—	—
Yazio; 7.10.8; August 8, 2022; 2014	21.5	18	8.2	7.4	4.6	—	—	4.1	—
LIFESUM; 11.0.0; June 30, 2022; 2011	20.5	14	7.2	6.8	4.5	—	—	—	—
MyNetDiary; 8.1.1; August 12, 2022; 2010	18.5	22	6.2	6.8	4.6	79	—	—	2.5
Second Nature; 6.10.2; August 11, 2022; 2016	12	28	8	8.5	4.6	84	—	—	—
Freeletics Nutrition; 1.27.12; January 27, 2020; 2016	7.5	21	6.7	5.5	3.8	—	—	—	2.5
Apps included in the first two rounds only									
HealthifyMe; 18.7.1; August 1, 2022; 2013	21.5	13	2.5	—	4.4	—	—	—	2
LIFE Extend[e]; 5.8.3; August 9, 2022; 2019	20.5	24	5.5	—	4.1	74	—	—	—
Noom; 10.28.0; August 17, 2022; 2011	17.5	12	—	—	4.2	80	5	—	—
Fastic Fasting App; 1.105.0; August 16, 2022; 2019	14	11	—	—	4.7	70	—	—	—
Fatsecret; 9.12.5.2; May 25, 2022; 2007	11	10	—	—	4.7	—	4	4	—
Eat This Much-Meal Planner; 2.0.12; August 12, 2022; 2016	11	13	—	—	4.4	—	—	—	2
8fit Workout & Meal Planner; 22.4.0; May 9, 2022; 2014	10	14	—	—	4.3	75	—	—	2.5
DietBet; 18.0.0; August 4, 2022; 2014	8.5	6	—	—	4.7	—	—	—	2.5
MyFitnessPal; 22.15.0; August 3, 2022; 2010	6.5	5	—	—	4.4	47	3	3	—
Contador de calorías; 3.7.3; April 28, 2022; 2014	6	6	—	—	4.5	—	—	4.5	—
Apps included in the first round only									
GetFit-Daily Meal Planner [e]; 1.3.4; April 23, 2023; 2020	16.5	—	—	—	1	—	—	—	2.5
MyPlate[e]; 3.5.3(63); September 30, 2021; 2015	14	—	—	—	4.6	—	—	—	—
Uplyfe-Precision Nutrition[e]; 1.8.1; September 21, 2021; 2020	13.5	—	—	—	4.2	70	—	—	—

[a]QUEST: Quality Evaluation Scoring Tool.

[b]ORCHA: Organization for the Review of Care and Health Applications.

[c]GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service).

[d]Not applicable.

[e]Became unavailable during the evaluation process.

Evaluation Round 1

Table S2 in Multimedia Appendix 1 displays the score of each QUEST item in the first round of evaluations, categorized by reviewer, with apps ranked according to their mean total score. We explored potential correlations between the QUEST scores, Play Store ratings, and the scores of each certification body and found no significant correlations among any combination (data not shown). Agreement reached between reviewers was rather low, with most κ coefficient values below 0.6 and the intercorrelation coefficient between 0.50 and 0.67 (see Table S3 in Multimedia Appendix 1). The main reasons for discrepancy were attributed to the difficulty in finding the information needed to complete the QUEST scale.

Evaluation Round 2

Four additional experts were asked to update the app review on an Android device and were given a maximum of 45 minutes per app to find the information. At the time of this evaluation, 3 (15.8%) apps (*MyPlate*, *GetFit-Daily Meal Planner*, and *Uplyfe-Precision Nutrition*) were no longer available, and the remaining 16 (84.2%) apps were scored (see Table 3). One of the reviewers used an iOS device (Apple iPhone) due to unavailable Android devices. The 7 (43.8%) apps that received a score above 20 (out of a maximum of 28) either in the first evaluation or in the update were selected for further assessment.

Challenges were encountered assessing the items "authorship," "attribution," and "tone." Some apps (particularly *LIFESUM*, *Freeletics Nutrition*, *Eat This Much-Meal Planner*, *LIFE Extend*, and *Second Nature*) lacked clear indications regarding authors and sources, necessitating external web searches to retrieve these data, and even then, this took considerable time. Additionally, reviewers noted that the "attribution" item could be easily altered to achieve higher scores; for example, referencing a source, even if it was a low-quality source or irrelevant to the app's health information, could lead to a high rating, as long as it was a highly scored type of study. Subjectivity was also a significant factor in evaluating the "tone" item, leading to disparities in scores. Finally, some of the discrepancies were explained by the differences in app versions (eg, *Freeletics Nutrition*, *Fresh Well*, and *Second Nature* improved considerably between review rounds, coinciding with version changes, whereas the opposite was true for *HealthifyMe*). In the case of *HealthifyMe*, this (and the low subjective score) led to its exclusion from the third round, despite scoring above 20 points in the first round. Discussion about the subjective scores and the contribution by the reviewer scoring all the selected apps in this round led to a more structured scoring procedure in the third round.

Evaluation Round 3

All experts participating in the previous rounds of evaluation, except one who was no longer available for the task, took part in this round. At the time of this evaluation (early 2024), LIFE Extend was not available, so only 6 (31.6%) apps were included in the third round. Discrepancies in the scores were discussed among researchers.

Subjective scores were also given to the selected apps by all six reviewers (see Table S4 in Multimedia Appendix 1 and Table 3). The scores were discussed in a joint meeting, which led to the selection of the top 5 (8.3%) apps: *Second Nature*, *Freshwell*, *Yazio*,*LIFESUM*, and *MyNetDiary*. The scores given by each evaluator are included in Multimedia Appendix 1.

Second Nature was consistently the most highly scored app and is also endorsed by the National Health Service (NHS) of the United Kingdom. Its present version includes clear references to the evidence supporting its contents. Usability is good and improved by the inclusion in the update of interactive content, such as videos with advice and encouragement. There is a wide variety of dietary patterns to choose from, and recipes with visual backup are provided. Furthermore, most of the features are free to use. However, food registering is rather cumbersome, and the tone of the recommendations could be more cautious.

Freshwell was developed by two British general practitioners and is used by the NHS in the United Kingdom. Usability is good, with pictures, recipes, and explanations. Long-term goals are included, promoting behavior change. A classification of foods is included based on the type of dietary pattern that is promoted (low carb). Study references are provided to support this recommendation, and a disclaimer is provided stating that the app provides educational content and is not to be considered medical advice. People with chronic conditions are advised to ask their health care provider. Other health-promoting pieces of advice include reducing sugar intake, snacks, and alcohol. The subjective score penalized the low-carb type of dietary pattern on which the app is based, since reviewers considered that there is not sufficient evidence to promote it to the general public as the "default" dietary pattern [25].

Yazio has high usability due to likable aesthetics, ease of use, accessibility, and the challenges included in the app. The challenges, which promote behavior change, can be chosen by the user from a list, including quitting chocolate, sugar, sweets, fast food, etc. The app includes information about healthy eating, which is supported by scientific evidence, and the standard recommended diet is Mediterranean style. However, it also includes dietary patterns with less scientific evidence of benefit, such as intermittent fasting or the keto diet. In fact, intermittent fasting is a prominent feature of this app, including challenges on how long you can fast for. A disclaimer recommends not using this pattern in the long term and seeing a physician if that is the intention of the user.

LIFESUM has a visually pleasant interface and agile and intuitive navigation, both of which lead to good user experience. Different objectives can be set, and the app includes easy and

advanced recipes, which help with organization and motivation for healthier eating, although all this is only available in the premium version. The app includes a score for easy comparison and progress tracking, but the criteria for the score are not transparent. Studies supporting the contents were not easily found and were not there for all the recommendations. Once again, the latter included intermittent fasting and keto diets. Minor bugs were identified in the form of nonfunctioning links.

MyNetDiary offers an attractive design, easy navigation, and a variety of resources to engage users. It provides information about different types of diets and highlights their key aspects. Examples of recipes for each type of diet are included with detailed descriptions. The content is supported by a trained specialist in the nutritional management of diabetes and other health problems. A diary function to track daily food intake is included, and data can be entered via text, audio, or barcode scanning. The premium version offers daily advice and feedback. Users can also access a nutrition blog written by the same specialist behind the app. Professionals can subscribe to connect with users interested in weight loss.

Discussion

Principal Findings

In this paper, we highlighted the need for a common evaluation framework to assess the quality of health apps and proposed a robust approach to this end. Our use case focused on apps promoting healthy eating. A total of 19 apps were included, which had already been positively assessed by dedicated certification bodies. The latter used a variety of evaluation criteria, and most did not include content quality among them. Following a rigorous, three-step evaluation process by seven experts, only 5 apps were recommended. The selection was based on a comprehensive assessment of the quality of their contents (using QUEST), as well as their usability and the use of behavior change techniques.

Health apps have emerged as promising tools to promote healthier lifestyles, due to their widespread accessibility, user friendliness, and cost-effectiveness [26-28]. However, the app market's diversity and the lack of systematic evaluation processes leave consumers with limited tools to judge which apps are effective, safe, and suited to their needs.

To address this issue, several national and international actions have attempted to assess health app quality and safety [17,20-23]. Our investigation identified 28 initiatives dedicated to the evaluation of health apps, 5 including apps fulfilling our inclusion criteria. Although each of these initiatives uses distinct criteria for app evaluation, they all generally emphasize similar key factors, such as data security, privacy, ease of use, accessibility, usability, support for user-health care professional communication, personalization options, and capacity to induce behavior change. In a recent review, the importance of adherence to scientific evidence, as well as additional features (eg, gamification and cocreation of the app with health professionals and users), were supported [29]. Although a multitude of tools and scores have been proposed to evaluate different aspects of app quality, as of today, one single robust procedure or set of criteria does not exist [30-32], although some national and international evaluation efforts have been developed [33-35]. Additionally, all the certification bodies reviewed in this study have cautioned that they cannot guarantee the precision, quality, trustworthiness, and effectiveness of the health apps they assess, increasing uncertainty for the users.

Acknowledging these challenges, our study sought to evaluate nutrition-focused health apps using a tool aimed at the health information within the apps, assuming that health information based on evidence would support safety and effectiveness. We selected QUEST to assess the apps, since it has been specifically designed to evaluate health information and incorporates items such as attribution or study type to evaluate the quality of information sources. Additionally, it assesses items such as conflict of interest and tone to evaluate how information is presented to the user, where any sort of advertising is penalized. We found no significant correlations among the Play Store scores for the apps, the QUEST total mean score, or any of the certifier scores, probably because they assess different dimensions. QUEST is a robust tool for the assessment of content quality but does not address usability, engagement, or behavior change techniques, all factors known to influence the perceived quality of an app [11]. Therefore, the research team decided to complement this assessment with a subjective score, which also included these aspects. Such factors may influence the app store ratings, and many are directly evaluated by the certification bodies but are overlooked by QUEST. In contrast, Play Store scores are based on user opinions, which can be influenced not only by all these factors but also by popularity, endorsement by influencers, aesthetics, or alignment with their own nutritional preconceptions, which might not be backed by scientific evidence.

Furthermore, it is important to acknowledge that apps are dynamic and subject to constant updates. Although the latter can be beneficial if they enhance the quality of app contents, they also pose a challenge for app evaluation, as acknowledged by a recent consensus [36]. Thus, continuous reassessment would be necessary to account for the changes. Access to older app versions may not always be possible, and certain apps may become completely unavailable over time. Any app marketed as a health app should be subject to special scrutiny, particularly regarding collection of sensitive data and potential for harmful effects on users. Previously proposed improvements include establishing national lists of tested and trusted health apps, creating a catalog of health apps accessible to patients only if prescribed by professionals and guidelines for app developers toward evidence-based, unbiased, high-quality apps, with wider assessment requirements for higher-risk tools [14,33,36-39].

Strengths and Limitations

Our work has several strengths that contribute to the field of health app evaluation. First, the adoption of QUEST to assess all the apps constitutes a rigorous approach for the evaluation of content quality. This systematic assessment of the content was complemented by a subjective evaluation by the panel of usability and behavior change techniques. Additionally, the study's methodological approach, involving multiple, independent reviewers and the evaluation of a diverse range of

apps, strengthens the validity and comprehensiveness of the results. Finally, the study's alignment with the existing literature and national and international certification/assessment bodies highlights its relevance and potential impact on improving health app quality and user safety.

We acknowledge that the study also has some limitations. The small sample size of apps reviewed and their focus on nutrition-related interventions may limit the generalizability of our results, since they represent an incomplete picture of the overall quality and safety of available health apps. The decision to exclude apps that solely function as food diaries and calorie counters may be controversial. We chose apps offering recommendations, meal plans, and recipes, because we focused on the quality of the information provided. Nevertheless, we acknowledge that food diaries and calorie counting could also help the users become more aware of their eating habits and induce a change in behavior despite the lack of any other specific nutritional advice and therefore be considered a type of intervention. Another limitation is the subjective nature (eg, tone) or the scoring difficulty (eg, attribution) of some of the items in QUEST. Discrepancies in the first round of evaluations, reflected by rather low κ coefficient values, were mainly attributed to the fact that the information to score some items might be difficult to find in the app or might even be found behind a link and thus not be obvious. Hence, in the second round of evaluations using QUEST, we standardized the time spent on the assessment of each app. In addition, some aspects of app quality are overlooked by QUEST (eg, usability, engagement, and capacity to induce positive behavior changes). Thus, QUEST is not a comprehensive enough tool to judge app quality on its own, but it is complementary to other measures. Finally, to overcome these difficulties, several rounds of revision were made, which made the evaluation process longer than initially planned.

To summarize, during this review, we encountered several challenges. In the first place, numerous certification bodies were scanned for evaluated apps. Creating a common repository collecting their criteria and decisions would facilitate this task [40]. In addition, the rapid change in mHealth apps poses a challenge to the timeliness of evaluation. This might be mitigated by the development of living systematic reviews, which are continuously updated [39] or the use of rapid reviews [41] and evidence maps [42]. Lastly, there is currently no standardized procedure to evaluate app quality. To address this, we included what we considered a critical and often overlooked dimension in previous evaluation efforts: a thorough assessment of the quality of the health information provided by the apps using QUEST scoring.

Conclusion

In conclusion, despite their previous evaluations by various certification bodies, only 5 of 19 apps promoting healthy eating met the quality standards set by our experts. Our study calls for enhanced scrutiny and regulatory measures to ensure that health apps meet rigorous standards of accuracy, reliability, and user safety. Guidelines for developers for the design of evidence-based, unbiased, high-quality health apps, as well as technological solutions for real-time monitoring of the apps, would address these challenges and improve reliability.

Acknowledgments

We thank Cristina Ruano Rodríguez (University of Las Palmas de Gran Canaria) for participating in the first round of evaluations. We would also like to thank Lilisbeth Perestelo (Servicio de Evaluación, Servicio Canario de la Salud) for fruitful methodological discussions. Where authors are identified as personnel of the International Agency for Research on Cancer or the World Health Organization (WHO), the authors alone are responsible for the views expressed in this paper, and the views do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer or WHO. This study was performed within WARIFA "(Watching the Risk Factors: Artificial Intelligence and the Personalised Prevention and Management of Chronic Conditions), funded by the European Union's Horizon 2020 research and innovation program (grant agreement 101017385). The sponsor played no role in the design; the collection, analysis, and interpretation of data; or the decision to publish the results. During the study, AMW received funding from the Instituto de Salud Carlos III (PI20/00846, PMP21/00069, and DTS 19/00067), the Fundación Mapfre Canarias 2023, the Sociedad Española de Diabetes, and the Fundación Canaria del Instituto de Investigación Sanitaria de Canarias (PIFIISC20/16 and PIFIISC22/08).

Data Availability

All relevant data are available for direct review within the paper.

Authors' Contributions

AD-G, GZZ, and AMW designed the study and summarized and analyzed the results. AD-G and GZZ performed the search and first round of assessments and drafted the manuscript. ITG, GS, AT-C, and MLA-M performed the second round of assessments. All authors participated in the discussions and final ratings of the apps, participated in the revision of the paper, and approved the final version. During the preparation of this work, the authors did not use any artificial intelligence tools/services at any time. The authors are responsible for the entire content of this publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Initial search strategy, list of certifying bodies examined, scores given to apps by evaluators, interobserver concordance analysis, and subjective assessments by reviewers.
[DOCX File , 60 KB - mhealth_v14i1e68737_app1.docx]

Multimedia Appendix 2
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[PDF File (Adobe PDF File), 86 KB - mhealth_v14i1e68737_app2.pdf]

References

1. Global digital health business outlook survey – 2022. Research2Guidance. 2022. URL: https://research2guidance.com/product/global-digital-health-business-outlook-survey-2022/ [accessed 2026-01-14]
2. Descargas mundiales de aplicaciones (apps móviles). Statista. URL: https://es.statista.com/estadisticas/574024/numero-de-descargas-mundiales-de-apps-mundo/ [accessed 2024-05-19]
3. Tuberculosis: global task force on digital health. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/tuberculosis-global-task-force-on-digital-health [accessed 2024-05-19]
4. Peterson CB, Hamilton C, Hasvold P, World Health Organization, Regional Office for Europe. From innovation to implementation: eHealth in the WHO European region. 2016 Sep 30. URL: https://www.who.int/europe/publications/i/item/9789289051378 [accessed 2026-01-14]
5. 325,000 mobile health apps available in 2017 – Android now the leading mHealth platform. Research 2Guidance. 2017. URL: https://research2guidance.com/325000-mobile-health-apps-available-in-2017/ [accessed 2026-01-14]
6. Daraz L, Morrow A, Ponce O, Beuschel B, Farah M, Katabi A, et al. Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the internet. J Gen Intern Med 2019 Sep;34(9):1884-1891 [FREE Full text] [doi: 10.1007/s11606-019-05109-0] [Medline: 31228051]
7. Guidance - MDCG endorsed documents and other guidance. European Commission. URL: https://health.ec.europa.eu/medical-devices-sector/new-regulations/guidance-mdcg-endorsed-documents-and-other-guidance_en [accessed 2026-01-14]
8. Manual on borderline and classification in the community regulatory framework for medical devices. Version 1.22. European Commission. URL: https://health.ec.europa.eu/system/files/2020-08/md_borderline_manual_05_2019_en_0.pdf [accessed 2026-01-14]
9. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies - developing and maintaining an innovative evidence framework with global impact. Digit Health 2021;7:20552076211018617 [FREE Full text] [doi: 10.1177/20552076211018617] [Medline: 34249371]
10. Unsworth H, Wolfram V, Dillon B, Salmon M, Greaves F, Liu X, et al. Building an evidence standards framework for artificial intelligence-enabled digital health technologies. Lancet Digital Health 2022 Apr;4(4):e216-e217 [FREE Full text] [doi: 10.1016/s2589-7500(22)00030-9]
11. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, Zamora-Zamorano G, Castro-Fernandez M, Alberiche Ruano MDP, WARIFA Consortium. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 04;25:e44030 [FREE Full text] [doi: 10.2196/44030] [Medline: 37140973]
12. Martín Fernández A, Marco Cuenca G, Salvador Oliván JA. Evaluación y acreditación de las aplicaciones móviles relacionadas con la salud. Rev Esp Salud Pública 2020;94(11 de agosto):e1-e11 [FREE Full text]
13. Assessment frameworks. European mHealth Hub. URL: https://web.archive.org/web/20211207214114/https://mhealth-hub.org/assessment-frameworks [accessed 2026-01-14]
14. Health apps repositories in Europe Internet. European mHealth Hub. URL: https://web.archive.org/web/20211024211815/https://mhealth-hub.org/health-apps-repositories-in-europe [accessed 2026-01-14]
15. 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]
16. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
17. Healthy living apps. VicHealth. URL: https://web.archive.org/web/20211116162051/https://www.vichealth.vic.gov.au/media-and-resources/vichealth-apps [accessed 2026-01-14]
18. Terhorst Y, Philippi P, Sander L, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
19. McKay F, 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]

20. my health apps. myhealthapps. URL: https://web.archive.org/web/20210725085246/https://myhealthapps.net/ [accessed 2026-01-14]
21. Delivering safe digital health. ORCHA. URL: https://appfinder.orcha.co.uk/about [accessed 2026-01-14]
22. The best apps for your health!. GGD AppStore. URL: https://web.archive.org/web/20211203144501/https://www.ggdappstore.nl/Appstore/Homepage/Sessie,Medewerker,Button [accessed 2026-01-14]
23. App library. Health Navigator. URL: https://web.archive.org/web/20211204173901/https://www.healthnavigator.org.nz/apps/ [accessed 2026-01-14]
24. Robillard J, Jun J, Lai J, Feng T. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inform Decis Mak 2018 Oct 19;18(1):87 [FREE Full text] [doi: 10.1186/s12911-018-0668-9] [Medline: 30340488]
25. Hassapidou M, Vlassopoulos A, Kalliostra M, Govers E, Mulrooney H, Ells L, et al. European Association for the Study of Obesity position statement on medical nutrition therapy for the management of overweight and obesity in adults developed in collaboration with the European Federation of the Associations of Dietitians. Obes Facts 2023 Jan 15;16(1):11-28 [FREE Full text] [doi: 10.1159/000528083] [Medline: 36521448]
26. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
27. Paramastri R, Pratama S, Ho D, Purnamasari S, Mohammed A, Galvin C, et al. Use of mobile applications to improve nutrition behaviour: a systematic review. Comput Methods Programs Biomed 2020 Aug;192:105459 [FREE Full text] [doi: 10.1016/j.cmpb.2020.105459] [Medline: 32234632]
28. Villinger K, Wahl D, Boeing H, Schupp H, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. Obes Rev 2019 Oct;20(10):1465-1484 [FREE Full text] [doi: 10.1111/obr.12903] [Medline: 31353783]
29. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos P, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
30. Health information quality assessment tools. OpenMD. URL: https://openmd.com/guide/assessing-health-information-quality [accessed 2026-01-14]
31. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021 Mar 19;11(3):e047001 [FREE Full text] [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
32. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli N, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. Evid Based Ment Health 2019 Feb;22(1):4-9 [FREE Full text] [doi: 10.1136/ebmental-2018-300069] [Medline: 30635262]
33. Segur-Ferrer J, Moltó-Puigmartí C, Pastells-Peiró R, Vivanco-Hidalgo RM. Line of methodological developments of the Spanish Network of Health Technology Assessment Agencies and National Health System Services. In: Health Technology Assessment Framework: Adaptation for Digital Health Technology Assessment. Madrid; Barcelona: Ministry of Health; Agency for Health Quality and Assessment of Catalonia; 2023.
34. Promoting a quality label for health apps!. Label2Enable. URL: https://label2enable.eu/ [accessed 2026-01-14]
35. Neal D, Engelsma T, Tan J, Craven M, Marcilly R, Peute L, et al. Limitations of the new ISO standard for health and wellness apps. Lancet Digital Health 2022 Feb;4(2):e80-e82 [FREE Full text] [doi: 10.1016/s2589-7500(21)00273-9]
36. Evidence standards framework for digital health technologies. National Institute for Health and Care Excellence (NICE). URL: https://www.nice.org.uk/what-nice-does/digital-health/evidence-standards-framework-esf-for-digital-health-technologies [accessed 2026-01-14]
37. van Haasteren A, Gille F, Fadda M, Vayena E. Development of the mHealth App Trustworthiness checklist. Digit Health 2019;5:2055207619886463 [FREE Full text] [doi: 10.1177/2055207619886463] [Medline: 31803490]
38. Dawson R, Felder T, Donevant S, McDonnell K, Card E, King C, et al. What makes a good health 'app'? Identifying the strengths and limitations of existing mobile application evaluation tools. Nurs Inq 2020 Apr;27(2):e12333 [FREE Full text] [doi: 10.1111/nin.12333] [Medline: 31854055]
39. Elliott J, Synnot A, Turner T, Simmonds M, Akl E, McDonald S, Living Systematic Review Network. Living systematic review: 1. Introduction-the why, what, when, and how. J Clin Epidemiol 2017 Nov;91:23-30 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.08.010] [Medline: 28912002]
40. Lopez-Alcalde J, Susan Wieland L, Barth J, Grainger R, Baxter N, Heron N, et al. Methodological challenges in systematic reviews of mHealth interventions: survey and consensus-based recommendations. Int J Med Inform 2024 Apr;184:105345 [FREE Full text] [doi: 10.1016/j.ijmedinf.2024.105345] [Medline: 38309237]
41. Garritty C, Gartlehner G, Nussbaumer-Streit B, King V, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021 Feb;130:13-22 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.10.007] [Medline: 33068715]
42. White H, Albers B, Gaarder M, Kornør H, Littell J, Marshall Z, et al. Guidance for producing a Campbell evidence and gap map. Campbell Syst Rev 2020 Dec;16(4):e1125 [FREE Full text] [doi: 10.1002/cl2.1125] [Medline: 37016607]

Abbreviations

ABACUS: App Behavior Change Scale
GGD: *Gemeentelijke Gezondheidsdienst* (Municipal Health Service)
MARS: Mobile App Rating Scale
mHealth: mobile health
NHS: National Health Service
ORCHA: Organization for the Review of Care and Health Applications
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUEST: Quality Evaluation Scoring Tool

Edited by L Buis; submitted 12.Nov.2024; peer-reviewed by M Agbede, S Mitra, B Eapen; comments to author 17.Apr.2025; revised version received 20.May.2025; accepted 14.Nov.2025; published 30.Jan.2026.

Please cite as:
Zamora Zamorano G, Déniz-García A, Torres-Castaño A, Álvarez-Malé ML, Gram IT, Skeie G, Wägner AM
The Landscape of Mobile Apps for Healthy Eating: Case Study for a Systematic Review and Quality Assessment
JMIR Mhealth Uhealth 2026;14:e68737
URL: https://mhealth.jmir.org/2026/1/e68737
doi:10.2196/68737
PMID:

Original Paper

Determining Cluster-Specific Differences in the Number of Days Required to Reliably Predict Habitual Physical Activity: Intraclass Correlation Resampling Analysis

Conor Jordan Murphy[1,2], PhD; Gabriel M Jouan[1,3], PhD; Katrin Y Friðgeirsdóttir[1,2], PhD; Anna Sigridur Islind[1,3], PhD; Jose M Saavedra[1,2], PhD; María Óskarsdóttir[3], PhD; Erna Sif Arnardóttir[1], PhD

[1]School of Technology, Reykjavik University Sleep Institute, Reykjavík University, Reykjavik, Capital Region, Iceland

[2]Sports Science Department, Physical Activity, Physical Education, Sport and Health (PAPESH) Research Centre, School of Social Sciences, Reykjavík University, Reykjavik, Capital Region, Iceland

[3]Department of Computer Science, Reykjavík University, Reykjavik, Capital Region, Iceland

Corresponding Author:
Conor Jordan Murphy, PhD
School of Technology, Reykjavik University Sleep Institute
Reykjavík University
Menntavegur 1
Reykjavik, Capital Region, 101
Iceland
Phone: 354 599 6200
Email: conorm@ru.is

Abstract

Background: Previous research has attempted to determine the minimum number of days of accelerometry required to reliably reflect an individual's physical activity. However, human behaviors on a day-to-day basis can be highly variable. As a consequence, the number of days required to reliably predict habitual physical activity is dependent on the variability that exists within an individual. There is a concern that adopting generic recommendations from previous research could provide unreliable estimates by failing to represent individuals with specific physical activity patterns.

Objective: The main aim of this study was to identify clusters of individuals with distinct physical activity patterns and to determine if the number of days of accelerometry data required to reliably estimate short- (7 days) and medium-term (28 days) physical activity differed between each unique cluster.

Methods: Accelerometry data were retrieved from 2 independent research studies. Participants during each study had their physical activity recorded using a Withings Scanwatch (Withings Health Solutions). Following a data eligibility process, agglomerative hierarchical clustering was used to identify clusters of individuals based on their physical activity. The clusters were determined using 4 dimensions; mean, SD, skewness, and kurtosis of the step count data. Intraclass correlation coefficients (ICCs) of step count were then calculated within each physical activity cluster. A series of ICCs were computed by separately comparing the average step count across the full periods (7 and 28, for the short- and medium-term analysis, respectively) to a series of averaged subsamples (ranging from 1-6 days and 1-27 days, for the short- and medium-term analysis, respectively). For each subsample, 500 random combinations were generated and compared, providing a distribution of ICCs for each subsample. An ICC of ≥0.80 identified when the subsample of days was sufficient to achieve appropriate reliability.

Results: Of 258 participant datasets, 149 were eligible for the short-term analysis and 64 were eligible for the medium-term analysis. Following agglomerative hierarchical clustering, 4 and 3 clusters of sufficient size (n≥12) were identified in the short-term and medium-term analyses, respectively. When considering the short-term analysis, to achieve a mean ICC score greater than or equal to 0.80, using all randomized combinations, the number of days ranged from 2 to 6 days depending on the physical activity cluster. For the medium-term analysis, the number of days required to achieve a mean ICC score greater than or equal to 0.80 ranged from 6 to 11 days. The short-term analysis clusters displayed more diversity in physical activity patterns than the medium-term analysis.

Conclusions: Physical activity patterns influence the number of days required to estimate habitual physical activity. Thus, to avoid unreliable estimates of physical activity, which could significantly impact the interpretation of results, researchers should be mindful of the physical activity patterns of their sample before adopting generic recommendations.

(JMIR Mhealth Uhealth 2026;14:e64323) doi:10.2196/64323

KEYWORDS

physical activity; accelerometry; reliability; intraclass correlation coefficient; human behavior; data; wearable devices; wearables; self-reporting; habitual; exercise; step counts; steps; walking

Introduction

Accelerometers are wearable devices that measure the acceleration of the body region to which the device is secured. Commercially, the signals are then processed and transformed into physical activity metrics that are relatable to the consumer, such as daily step count. The popularity of accelerometry as a method to quantify physical activity is growing due to the objectivity and availability of such devices. Importantly, the use of accelerometry provides an alternative option to self-reporting questionnaires, which can lack strong reliability and validity [1].

The field is rapidly developing with the advancement of new technology and the commercialization of devices. As a consequence, the approaches used to interpret collected data can vary and lack a consensus [2]. An area of importance relates to the minimum number of valid days required to reliably reflect a person's habitual level of physical activity [3]. Unfortunately, many studies with important implications, such as those linking physical activity to health outcomes, have used varying criteria to define a participant's habitual level of physical activity. For example, the United Kingdom biobank guidelines advise researchers to exclude participants from their accelerometry dataset who have fewer than 3 valid days of data [4], which has been adopted in many impactful studies [5-7]. In contrast, the National Center for Health Statistics, which administers the National Health and Nutrition Examination Survey in the United States, does not advise on the number of valid days that should be available to validate inclusion when using their accelerometry datasets. As a consequence, researchers have adopted various inclusion criteria ranging from 1 to 5 days [8-10].

From a research perspective, the fewer days required could help relieve the study burden on the participant and better support the use of study resources. Past reviews have suggested 4 days would suffice to achieve sufficient reliability [2,11]. Furthermore, a more recent comprehensive study using a large Singaporean sample recommended that at least 3 and 5 measurement days of step count data were needed to predict weekly and monthly time windows, respectively [12]. Despite the existing recommendations, providing a consensus for the minimum number of days required to predict habitual physical activity can be challenging and fraught with error if the uniqueness of the sample in question is not considered. Human behaviors on a day-to-day basis can be highly variable; thus, the number of days required to reliably predict habitual physical activity is dependent on the variability that exists within an individual [13]. Intuitively, individuals with high consistency or limited variability in physical activity may be predicted with fewer days than an individual with low consistency and high variability. As a consequence, the identification of an optimal number of days could be problematic. Another source of consideration relates to the formula used to calculate reliability. In the physical activity literature, the intraclass correlation coefficient (ICC) is commonly used to grade the reliability of physical activity and a common standard used to indicate acceptable reliability is an ICC of 0.8 [13,14]. However, broadly, this formula approximates the within-subject variability by comparing its magnitude to the between-subject variability. Thus, an increase in diversity between individuals will inherently reduce the reliability score without any change in variability within individuals. This situation poses a potential concern that predictions from large heterogenic samples may not translate to groups of individuals with specific physical activity patterns.

Therefore, the main aim of this study was to identify clusters of individuals with distinct physical activity patterns and to determine if the number of days of accelerometry data required to reliably estimate short- (7 days) and medium-term (28 days) physical activity differed between each unique cluster. This study used clustering analysis to allocate individuals from a heterogenic population into distinct physical activity clusters. Subsequently, reliability estimates were derived for each cluster and compared. The hypothesis was that the number of days to reliably predict (1) short- and (2) medium-term physical activity differs between clusters.

Methods

Study Design

This is a secondary, longitudinal analysis using combined data from 2 independent studies. Only the objectively measured physical activity data (daily step counts) and demographic descriptors were used from each dataset. All other variables from the original studies were excluded from the current analysis. The analysis was conducted at Reykjavik University in Iceland as part of the European Union Horizon 2020-funded Sleep Revolution research project [15].

Setting

Study 1 was an observational study conducted at the Reykjavik University Sleep Institute, aimed at exploring the associations between physical activity and markers of obstructive sleep apnea (OSA) severity [16]. A total of 66 adults were recruited, representing a broad range of sleepers, including healthy individuals, snorers, and those with suspected or diagnosed OSA. Each participant underwent a 3-night, self-applied somnography study at home, along with assessments of anthropometry, body composition, and both subjective and objective physical activity. Importantly, for the purpose of this analysis, objective physical activity was measured using a smartwatch device (Withings Scanwatch, Withings Health Solutions) worn on the nondominant wrist for a 3-month period.

Study 2 was a 12-week, 3-arm randomized controlled trial conducted at the Reykjavik University Sleep Institute [17]. The study was registered in ISRCTN 16974764. It aimed to evaluate

the effects of an exercise program and a lifestyle app on OSA, physical health, and quality of life in adults with mild-to-moderate OSA or habitual snoring. A total of 192 eligible participants were randomized into exercise, app, or control groups. Randomization was stratified using an algorithm to ensure balance across age, gender, BMI, and apnea-hypopnea index (AHI). Participants completed baseline assessments, including a one-night type 3 sleep study, body composition, and physical activity measures. The exercise group attended structured sessions 3 times per week, and the app group engaged in daily behavioral tasks via a health app (Sidekick Health, Reykjavík). Objective physical activity was assessed via a smartwatch device (Withings Scanwatch, Withings Health Solutions) worn on the nondominant wrist over the 12-week study period.

Participants

Participants from each study, who are described in the subsequent paragraphs, were pooled to explore habitual physical activity patterns. Study 1 recruited a heterogeneous sample from the general adult population (≥18 years) with the goal of reflecting natural variation in age, sex, and body morphology. Participants were not excluded based on sleep health status. Study 2 was a randomized controlled trial targeting adults aged 18–50 years of age that were categorized as overweight or obese (BMI ≥25<42 kg/m²), who were physically inactive and had either mild-to-moderate OSA (AHI ≥5 and <30) or habitual snoring (≥10%). Shift workers and individuals undergoing OSA treatment were excluded. Eligible participants underwent a one-night type 3 sleep study to confirm OSA.

Variables

The primary variable analyzed was daily step count, measured objectively using accelerometry-based methodology. Step count served as the key indicator of physical activity throughout the study period.

Data Measurements

Demographics

Data were collected using a custom-developed questionnaire. REDCap (Research Electronic Data Capture) survey software version 9.3.1 (Vanderbilt University) was used to collect participant responses [18,19].

BMI

Height and weight (digital scale; TANITA MC-780, Tanita Corporation) were measured and used to calculate BMI.

Smartwatch

Participants were instructed to wear a smartwatch (Withings Scanwatch, Withings Health Solutions) for the study period. The smartwatch generated physiological and physical activity parameters that included step count, distance covered, elevation, sleep, heart rate, oxygen saturation, passive calories, and active calories. Due to the substantial correlations between the physical activity parameters, step count was selected as the parameter of primary interest. The generated data were available in two forms: (1) timestamped data and (2) daily aggregated data. Additional details of the smartwatch are described by [20].

Bias

Several potential sources of bias were considered in the study design and analysis. First, selection bias was likely present in Study 2, as participants were specifically recruited based on specific inclusion criteria around OSA. However, Study 1 reflected a broader segment of the general adult population and may have helped offset this limitation by introducing more heterogeneity in age, sex, BMI, and physical activity profiles. Nonetheless, the combined sample remains somewhat weighted toward individuals with OSA. Measurement bias was addressed by using objective step count data from smartwatches rather than self-reported physical activity, reducing recall and reporting biases. However, inaccuracies could arise from nonwear time or device limitations; to mitigate this, systematic methods were used to exclude days with insufficient wear time [2,21]. Similarly, to address potential bias from missing days, nonconsecutive valid days were allowed but only within specified windows described in more detail in the Sample Size section.

Sample Size

All participants from Study 1 and Study 2 were considered eligible for analysis, but their inclusion depended on passing specific wear-time criteria.

Valid Day

Given the longitudinal nature of the study, participants were not required to report nonwear time, which influenced the accuracy of step count data. Therefore, a systematic approach to identify nonwear periods was established. Gaps in the timestamped “calories earned” data exceeding 60 minutes were classified as nonwear periods. Summing the nonwear periods that occurred between consecutive sleep periods permitted the calculation of wear time during waking hours. In line with previous approaches and retaining statistical power [2,21], days with less than 10 hours of wear time during waking hours were considered insufficient to accurately estimate step count and excluded.

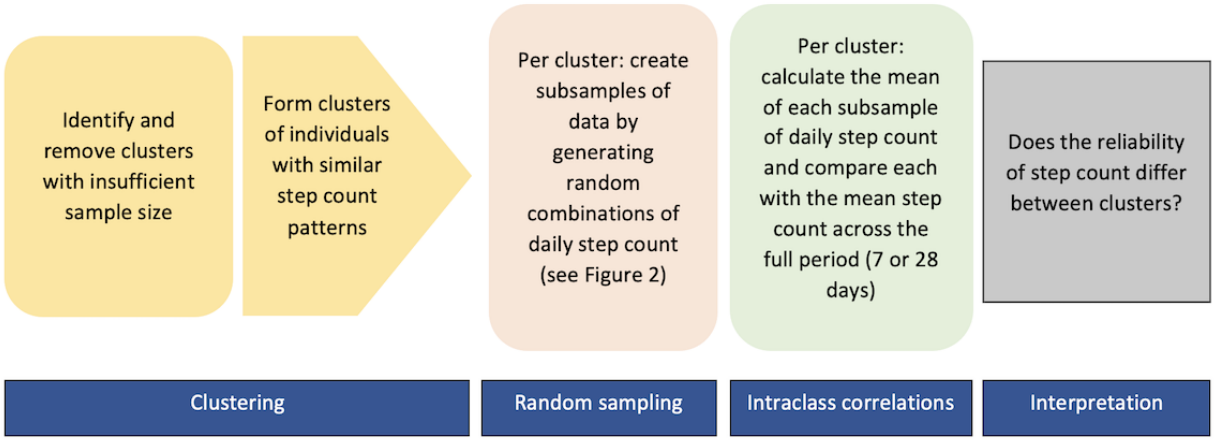
Valid Period

Using the step count data filtered for valid days, one period of 7 days and another of 28 days were selected for each participant in the short- and medium-term analyses, respectively. Periods with nonconsecutive days were allowed if there were (1) 7 valid days within a 10-day period in the short-term analysis or (2) 28 valid days within a 40-day period in the medium-term analysis. Periods with full consecutive sequences or with the fewest days skipped were preferentially chosen, but if more than one of these sequences existed, then the selection was randomized. Nonconsecutive days were permitted due to the difficulty of achieving full consecutive sequences of data, particularly when considering the medium-term analysis. For the inspection of data, the chosen periods of 10 and 40 days were chosen to allow us to retain a sufficient sample size without compromising the representation of physical activity for the given period.

Data Analysis of Quantitative Variables

The methodological approach to the analysis of the daily step count variable is displayed in a visual format in Figure 1. Using unsupervised machine learning, we identified clusters of individuals that shared more similar physical activity patterns than those in the other identified clusters. Agglomerative hierarchical clustering with complete-linkage, a distance-based algorithm, was used to identify clusters [22]. This form of clustering was chosen because it can be used to identify outliers and does not require a predetermined number of clusters. The clustering analysis was conducted for both short- and medium-term physical activity patterns, separately. The clustering analysis first involved running the algorithm to identify sufficiently sized clusters. Given that the calculated reliability metrics are influenced by between-subject variability, smaller clusters may be more prone to an underestimation or overestimation of reliability depending on participant similarities. Thus, if any clusters with less than 12 participants were identified, they were removed, and the remaining clusters were retained for the full analysis. The clusters were determined based on their similarities across 4 dimensions, those being the scaled and centered mean, SD, skewness, and kurtosis of the step count data. These dimensions represent the 4 moments of distribution, which quantitatively reflect each participant's physical activity patterns in the context of an average, day-to-day variability and range of values.

Figure 1. Methodological approach to study analysis.



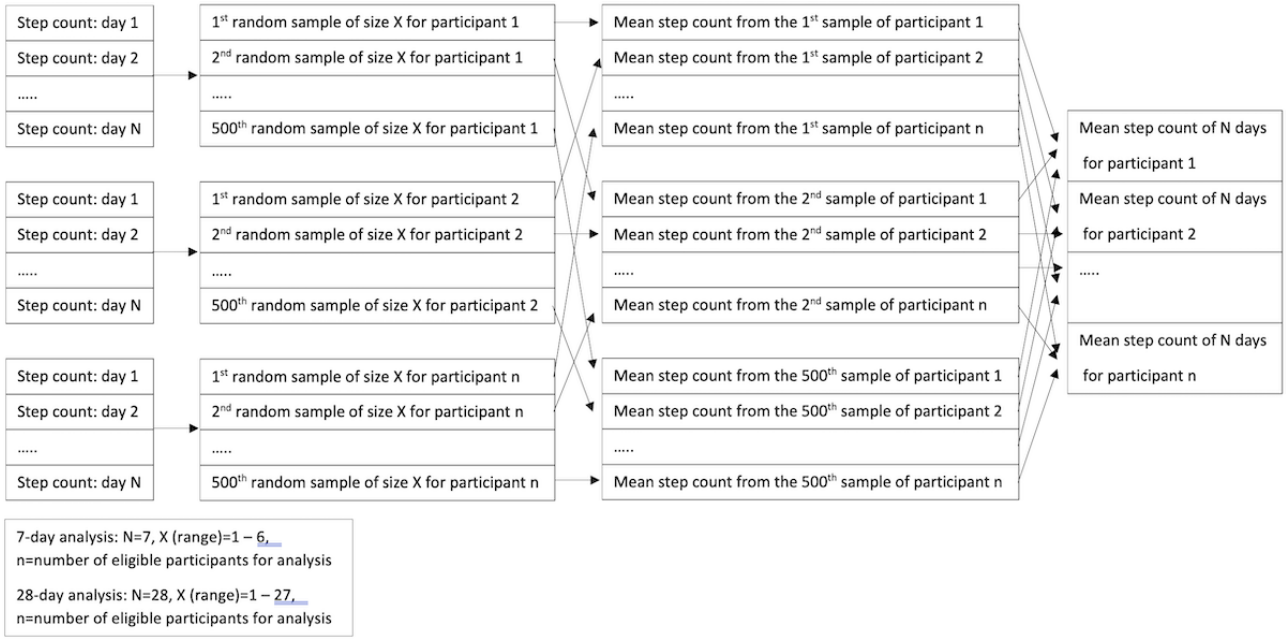
Statistical Analysis

Following the identification of clusters, the reliability of step count within each cluster was determined, along with a baseline comparison (unclustered data). ICCs were derived using linear mixed models with one-way random effects. A series of ICCs were computed by separately comparing the average step count across the full period (7 and 28 days for the short- and medium-term analysis, respectively) to a series of averaged subsamples (ranging from 1 to 6 days and 1 to 27 days for short- and medium-term analysis, respectively). As the subsamples could include any combination of days from the full period, for each participant, resampling methods without replacement were used to ensure an appropriate number of combinations were tested. For each subsample, 500 random combinations were generated and compared to the average step count across the full period (see Figure 2). Prior testing of the data showed that when 500 combinations were used for ICC predictions, the results remained stable when repeated. Thus, the range of ICC values possible for each subsample comparison was produced. In line with previous research, a threshold value of 0.80 was used to infer "acceptable" reliability [13,14].

It is necessary to point out the benefits of the Monte Carlo resampling approach used. This approach reduces the influence of nonnormality and heterogeneity of within-subject variances by using aggregated participant-level means and repeated ICC estimations. Given that ICCs were computed from mean step counts, assumptions of normality and homogeneity of variance are less probable, as justified by the Central Limit Theorem. Additionally, using 500 iterations per subset further enhances the robustness of the reliability estimates by reducing the influence of a typical combinations.

All the analyses were conducted in R (version 4.3.2; R Core Team). Specifically, the "ICC" package was used to extract the ICCs.

Figure 2. Data randomization of subsamples.



Ethical Considerations

Both studies received ethical approval from the National Bioethics Committee of Iceland and the Icelandic Data Protection Agency (21-170; 22-082). Participants provided written informed consent to participate, which also extended consent to secondary analysis. All data were stored anonymously and processed on a secure high-performance computing cluster to ensure efficient handling while maintaining data security and integrity. Finally, participants did not receive any compensation for their participation.

Results

Participants and Descriptive Data

Of the 258 original participants available from Study 1 and Study 2, a total of 149 participants possessed 7 valid days (>10 hour wear time during wake) across a 10-day period and were included in the short-term analysis, and 64 participants possessed 28 valid days across a 40-day period and were included in the medium-term analysis.

In the short-term analysis, 76 of the participants were male, and 73 were female. On average, the participants were 41 (SD 5) years of age and were categorized in the obese category (BMI: mean 31, SD 4). Age ranged from 19 to 76 years old, while BMI ranged from 19 to 42. In the medium-term analysis, half of the participants were male (n=37) and half of the participants were female (n=37). On average, the participants were 45 (SD 9) years of age and were categorized in the obese category (BMI: mean 31, SD 5). Age ranged from 23 to 65 years old, while BMI ranged from 20 to 41. The participants that were eligible for both analyses did not have any missing data.

Outcome Data

Clustering in Short-Term Analysis

Following agglomerative hierarchical clustering, 4 clusters of sufficient size (n≥12) were identified from a visual inspection of the dendrogram. The clusters were chosen by attempting to achieve a balance between appropriate sample sizes and within-cluster cohesion. The 4 dimensions of daily step count used to form each cluster are reported in Table 1.

In comparison to the other clusters, Cluster 1 displayed moderate physical activity levels and day-to-day variability. However, Cluster 1 had a highly skewed physical activity pattern and high kurtosis, suggesting these participants tended to have days that, when compared to their habitual physical activity levels, were considered extremely low or high. In contrast, the other clusters had kurtosis values below or slightly above 3, suggesting the number of days classed as outliers (extreme high or low physical activity) was similar to or less than that seen in a normal distribution pattern. Cluster 2 displayed the lowest relative variability in physical activity levels, while participants in Cluster 3 had the lowest overall physical activity levels, with habitual levels of 4391 (SD 1058) steps per day. Cluster 4 displayed the highest physical activity levels (mean 7179, SD 1028 steps per day), but also the highest day-to-day variability. Given the low kurtosis and moderate skewness, this would suggest the participants in Cluster 4 displayed a mixture of physical activity days.

Table 1. Dimensions of daily step count used to isolate clusters of individuals in the short-term analysis.

Short-term Cluster (n)	Mean steps, mean (SD)	SD steps, mean (SD)	Skewness, mean (SD)	Kurtosis, mean (SD)
1 (n=36)	5883 (1298)	2192 (727)	1.28 (0.48)	5.34 (1.08)
2 (n=22)	5980 (1538)	1619 (685)	0.82 (0.26)	3.67 (0.99)
3 (n=42)	4391 (1058)	1467 (586)	0.50 (0.35)	2.20 (0.83)
4 (n=30)	7179 (1028)	3068 (1132)	0.55 (0.37)	1.96 (0.70)

Clustering in Medium-Term Analysis

Following agglomerative hierarchical clustering, three clusters of sufficient size (n≥12) were retained by visually inspecting the dendrogram. The four dimensions of daily step count used to form each cluster are reported in Table 2.

Compared to the short-term analysis, the clusters within the medium-term analysis displayed less diversity. Modest day-to-day variability in physical activity was found in Cluster 1, but the participants in this cluster had the highest physical activity levels (mean 7671, SD 1052 steps per day). In contrast, participants in Cluster 2 displayed the lowest levels of physical activity (mean 4746, SD 947 steps per day) with the highest relative variability in physical activity. Cluster 3 displayed the least relative variability in physical activity and a low degree of skewness. This suggests that the physical activity levels for these participants were usually around the mean alongside an expected number of higher and lower physical activity days (normal distribution). The number of extremes in physical activity was small in all clusters and did not display distinctly different tails to a normal distribution pattern.

Table 2. Dimensions of daily step count used to isolate clusters of individuals in the medium-term analysis.

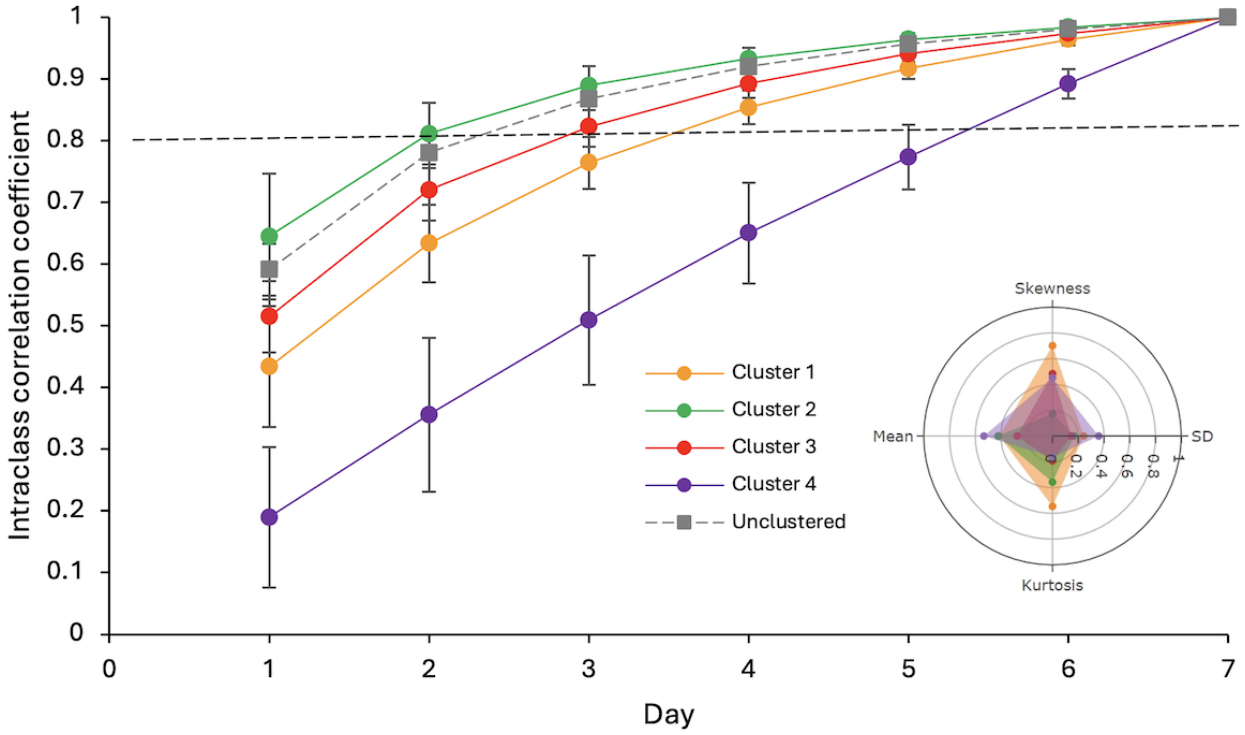
Medium-term Cluster (n)	Mean steps, mean (SD)	SD steps, mean (SD)	Skewness, mean (SD)	Kurtosis, mean (SD)
1 (n=23)	7671 (1052)	2955 (655)	0.66 (0.45)	3.65 (1.11)
2 (n=14)	4746 (947)	2265 (381)	0.87 (0.24)	3.31 (0.58)
3 (n=16)	5775 (988)	1764 (398)	0.31 (0.28)	2.87 (0.72)

Main Results

Intraclass Correlations in Short-Term Analysis

To achieve a mean ICC score above 0.80, calculated using all randomized combinations for each day subset, Cluster 2 required 2 days, Cluster 3 required 3 days, Cluster 1 required 4 days, and Cluster 4 required 6 days (Figure 3). The minimum number of days needed to achieve an ICC score above 0.80 in all 500 random combinations was 4 in Clusters 2 and 3, 5 in Cluster 1, and 7 in Cluster 4. As a baseline comparison, when the data were unclustered, 3 days were required to achieve a mean ICC score above 0.80 in all 500 random combinations.

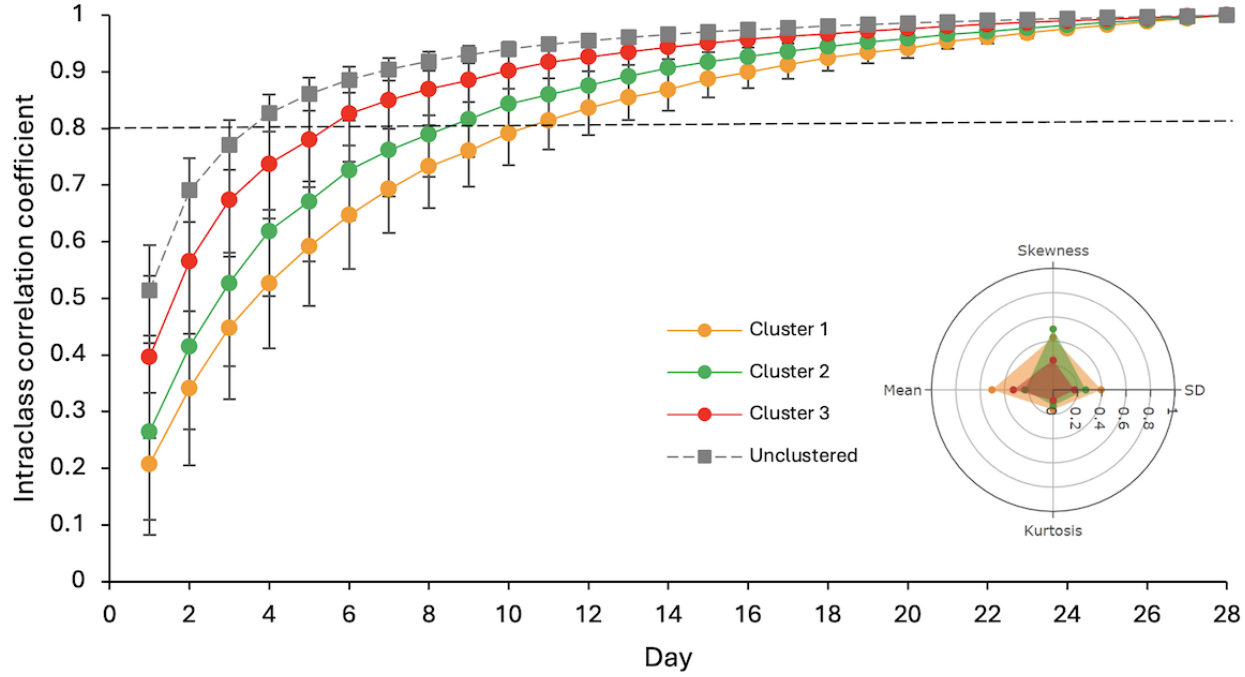
Figure 3. Range of intraclass correlation coefficients across all randomized combinations for each cluster from short-term analysis. Scores ≥ 0.80 are classed as "acceptable" agreement. Data are mean (SD).



Intraclass Correlations in Medium-Term Analysis

In order to achieve a mean ICC score greater than 0.80, calculated using all randomized combinations for each day subset, Cluster 3 required 6 days, Cluster 2 required 9 days, and Cluster 1 required 11 days (Figure 4). The minimum number of days needed to achieve an ICC score above 0.80 in all 500 random combinations was 11 in Cluster 3, 18 in Cluster 2, and 17 in Cluster 1. As a baseline comparison, when the data were unclustered, 4 days were required to achieve a mean ICC score above 0.80 in all 500 random combinations.

Figure 4. Range of intraclass correlation coefficients across all randomized combinations for each cluster from medium-term analysis. Scores ≥ 0.80 are classed as "acceptable" agreement. Data are mean (SD).



Discussion

Principal Findings

The main aim of this study was to identify clusters of individuals with distinct physical activity patterns and to determine if the number of days of accelerometry data required to reliably estimate short- (7 days) and medium-term (28 days) physical activity differed between each unique cluster. Clustering analysis confirmed that, in a heterogenic sample, clusters of individuals could be identified based on their physical activity patterns. In agreement with our hypothesis, the physical activity patterns of different individuals were highly relevant to the number of days needed to reliably predict both short- (7 days) and medium-term (28 days) physical activity, ranging from 2 to 6 days and 6 to 11 days, respectively, by cluster. Therefore, the minimum number of days should be assessed within a particular study cohort, instead of relying on generic recommendations.

One major consideration of using accelerometry for longitudinal physical activity research is that participant monitoring is a continuous process, which can require significant study resources and can challenge participant adherence. Therefore, having the flexibility to monitor less or lose more days, without compromising the validity of the data, is of significant benefit. However, the results from the current study show that, while this is possible, it can be erroneous to use a generic number of days across all participants. This supports the need to be cautious when considering a one-size-fits-all approach when it comes to collecting accelerometry data. For example, past reviews have suggested 4 days would suffice to achieve sufficient reliability in physical activity [2,11]. However, based on the results from the current study, this will be highly dependent on the physical activity patterns that exist within the specific research sample being studied. Another potential limitation of previous recommendations is the use of the Spearman-Brown prophecy formula to estimate reliability [23-25]. This statistical method depends on achieving homogeneity of variances, which may be unrealistic to achieve in many datasets. The resampling approach in this study helped address this problem, as it does not rely on this formula and is less sensitive to violations of homogeneity of variance. By repeatedly calculating ICCs from randomly selected combinations of days, the method reduces the influence of unequal within-subject variances on individual ICC estimates, resulting in a more robust overall estimate of reliability. Similarly, a nonclustered ICC may violate the assumption of homogeneity of variance if the physical activity patterns of the sample are diverse. Clustering can help group individuals with similar within-subject variance, improving the likelihood that model assumptions are satisfied.

The results from both the short- and medium-term analyses highlighted that clusters with relatively low variability in their physical activity required fewer days to reliably estimate physical activity. In contrast, clusters with high variability required significantly more days. Referring specifically to the data, depending on the cluster, the mean number of days required ranged from 2 to 6 for the short-term analysis and 6 to 11 for the medium-term analysis. This makes generalized recommendations as reported in previous research [26-28] difficult to provide. Interestingly, within this study's sample, the physical activity patterns of clusters in the medium-term analysis were more similar to each other than those observed in the short-term analysis. This implies reduced variability as participants settle into comparable daily activity patterns with larger measurement windows. Nonetheless, it is important to mention that, for one cluster, the minimum number of days needed to achieve sufficient reliability in all 500 random combinations was still noteworthy (17 days).

To an extent, these results agree with the work by [12] that recommended at least 3 and 5 measurements days of step count data were needed to predict weekly and monthly time windows, respectively. However, it highlights the risk of adopting the minimum limit of these recommendations for all participants. Instead, an approach that first attempts to understand the sample that is being studied could be beneficial. For example, if pilot testing is being conducted, this could provide an opportunity to collect prior accelerometer data on the group of participants. Depending on whether the physical activity patterns are expected to be similar or not, a cluster analysis could be used to determine the minimum number of days needed for each subset of participants. Alternatively, while physical activity questionnaires may provide an inexpensive and less-demanding method to group individuals, they can be prone to recall bias, as we have previously shown [16].

Practically, this research demonstrated that a heterogeneous sample can be divided into subgroups with distinct physical activity patterns. As a consequence, these clusters can influence the reliability of predicting habitual physical activity and, thus, the broader interpretation of the results. Previous research linking physical activity to health outcomes has estimated a participant's habitual level of physical activity using subsamples ranging from 1 to 5 days [8-10]. Unfortunately, the validity of such results is tied to the accuracy of this physical activity estimation. Researchers who do not fully consider the potential heterogeneity in their sample may be at potential risk of underestimating within-subject variability, which could potentially influence their interpretation of the results. Therefore, safeguards should be considered to improve physical activity estimates that include clustering, using appropriate ICC methods, and including complementary reliability metrics (standard error of measurement, coefficient of variation, etc).

Another point of consideration is that, in research studies, accelerometry is often used to quantify the rest-activity cycle of participants, which spans physical activity, sedentary time, and sleep. While this research focused on physical activity, the reliability of sedentary time and sleep will also display different levels of within-subject variability. Therefore, if researchers intend to quantify all 3 components of the rest-activity cycle, it is important to consider whether the number of days chosen is sufficient to capture the variability across all 3 components. For example, before individual sleep behaviors become stable, they require many more days of data than the general recommendation of 14 days [3]. However, more research exploring clustered samples is required because results from grouped data have reported differing results [29-31], which is likely a reflection of the level of heterogeneity across the sample in each analysis.

The current study is not without its limitations. First, the specific results of the current study reflect the discrete physical activity patterns of the analyzed sample. Nonetheless, the core concepts are universal and generalizable to other research samples in this area. With a larger sample size, more clusters may have been identified, or rather, a greater number of participants may have been placed within each identified cluster. For example, a number of clusters that were identified had insufficient sample sizes to conduct the analysis on and thus were excluded. It is plausible that the individuals with these excluded physical activity patterns may have required even more or fewer days to reliably estimate their physical activity over time, compared to those seen in the reported results. Importantly, while the physical activity patterns of the excluded individuals were uncommon in this study, they may better reflect physical activity patterns seen in a different study. Notably, while our sample was relatively heterogeneous, there was a higher percentage of individuals with large BMIs, which may have contributed to a small bias, as BMI has been shown to be associated with lower physical activity patterns [32]. Nonetheless, if the sample in this study had more variance, it likely would have strengthened the results while also improving the study's generalizability. In the current study, 10 or more hours of wear time during waking hours for a given day was considered sufficient for inclusion. Therefore, certain days with missing physical activity data will inherently be included and can influence reliability predictions [24]. Nonetheless, this decision was made to align with previous approaches [2,21], and to balance accuracy with the retention of information. Finally, ICCs are based on the proportion of between-to within-subject variance; thus, results from separate samples can differ based on within-subject variance (day-to-day variability of steps within a participant) but also, between-subject variance (variability of steps between participants). As alluded to in the introduction, this poses a potential concern that recommendations from large heterogeneous samples may misrepresent research samples that are more homogenous in nature. Unfortunately, while widely used, this is the limitation of using ICCs in the context of estimating physical activity. Importantly, our clustering analysis attempted to limit the between-subject variance by identifying clusters of individuals with similar physical activity patterns based on the 4 moments of distribution.

Conclusion

In conclusion, the number of days required to reliably estimate physical activity differs between clusters of individuals characterized by distinct physical activity patterns across both the short- and medium-term. To avoid unreliable estimates of physical activity, researchers should be mindful of the sample they are studying and how this may influence the minimum number of days required to reliably reflect physical activity. This study showcases that one-size, indeed, does not fit all when it comes to collecting accelerometry data.

Acknowledgments

The authors are grateful to all the participants for their involvement. The authors would like to thank the Sports Science Department for providing the Facility of Social Sciences at Reykjavík University, Reykjavík (Iceland) to carry out Study 2. We would also like to thank Birta Sóley, Steinunn Svansdóttir, Karitas Kjartansdóttir, and Katrín Gústafsdóttir for their work regarding recruiting participants, the preliminary Study 1 and Study 2. Also, the authors would like to thank Kristín Anna Ólafsdóttir for her extensive contribution to ethical applications and personnel management. We used the generative AI tool ChatGPT by OpenAI to support with code generation, which was reviewed and corrected by the study group, and to restructure writing.

Funding

The authors declare that financial support was received for the research, authorship, and/or publication of this article. The authors of this paper have received funding to perform this work and write this paper from the European Union's Horizon 2020 research and innovation program (grant agreement 965417). The Sleep Revolution project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 965417.

Authors' Contributions

Conceptualization was done by CJM, GMJ, JMS, MO, and ESA. Data curation and formal analysis were performed by CJM and GMJ. Funding acquisition was secured by JMS, ESA. Investigation was carried out by CJM and KYF. Methodology was developed by CJM, GMJ, JMS, MO, and ESA. Project administration was managed by CJM. Resources were provided by ASI. Software development was completed by CJM and GMJ. Supervision was provided by ASI, JMS, MO, and ESA. Validation was conducted by CJM and GMJ. Visualization was done by CJM and GMJ. Writing of the original draft was performed by CJM, while review and editing were carried out by CJM, GMJ, KYF, ASI, JMS, MO, and ESA.

Conflicts of Interest

The Sleep Revolution pays KYF PhD students' salaries and CJM, GMJ postdoctoral researchers' salaries. ESA reports honoraria from Nox Medical, ResMed, Jazz Pharmaceuticals, Linde Healthcare, Wink Sleep, Apnimed, and Vistor. A former member of Philips Sleep Medicine, Innovation Medical Advisory Board, and Lilly Medical Advisory Board. As well as currently being the President of the European Sleep Research Society. None for other authors.

References

1. van Poppel MNM, Chinapaw MJM, Mokkink LB, van Mechelen W, Terwee CB. Physical activity questionnaires for adults: a systematic review of measurement properties. Sports Med 2010;40(7):565-600. [doi: 10.2165/11531930-000000000-00000] [Medline: 20545381]
2. Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic review and practical considerations. Sports Med 2017;47(9):1821-1845 [FREE Full text] [doi: 10.1007/s40279-017-0716-0] [Medline: 28303543]
3. Óskarsdóttir M, Islind AS, August E, Arnardóttir ES, Patou F, Maier AM. Importance of getting enough sleep and daily activity data to assess variability: longitudinal observational study. JMIR Form Res 2022;6(2):e31807 [FREE Full text] [doi: 10.2196/31807] [Medline: 35191850]
4. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK biobank study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
5. Dempsey PC, Strain T, Khaw K, Wareham NJ, Brage S, Wijndaele K. Prospective associations of accelerometer-measured physical activity and sedentary time with incident cardiovascular disease, cancer, and all-cause mortality. Circulation 2020;141(13):1113-1115 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.119.043030] [Medline: 32223676]
6. Raichlen DA, Aslan DH, Sayre MK, Bharadwaj PK, Ally M, Maltagliati S, et al. Sedentary behavior and incident dementia among older adults. JAMA 2023;330(10):934-940 [FREE Full text] [doi: 10.1001/jama.2023.15231] [Medline: 37698563]
7. Cao Z, Min J, Chen H, Hou Y, Yang H, Si K, et al. Accelerometer-derived physical activity and mortality in individuals with type 2 diabetes. Nat Commun 2024;15(1):5164 [FREE Full text] [doi: 10.1038/s41467-024-49542-0] [Medline: 38886353]
8. Tudor-Locke C, Brashear MM, Johnson WD, Katzmarzyk PT. Accelerometer profiles of physical activity and inactivity in normal weight, overweight, and obese U.S. men and women. Int J Behav Nutr Phys Act 2010;7:60 [FREE Full text] [doi: 10.1186/1479-5868-7-60] [Medline: 20682057]
9. Shim J, Fleisch E, Barata F. Wearable-based accelerometer activity profile as digital biomarker of inflammation, biological age, and mortality using hierarchical clustering analysis in NHANES 2011-2014. Sci Rep 2023;13(1):9326 [FREE Full text] [doi: 10.1038/s41598-023-36062-y] [Medline: 37291134]
10. Sakal C, Li T, Li J, Li X. Predicting poor performance on cognitive tests among older adults using wearable device data and machine learning: a feasibility study. NPJ Aging 2024;10(1):56 [FREE Full text] [doi: 10.1038/s41514-024-00177-x] [Medline: 39587119]
11. Trost SG, McIver KL, Pate RR. Conducting accelerometer-based activity assessments in field-based research. Med Sci Sports Exerc 2005;37(11 Suppl):S531-S543. [doi: 10.1249/01.mss.0000185657.86065.98] [Medline: 16294116]
12. Yao J, Tan CS, Lim N, Tan J, Chen C, Müller-Riemenschneider F. Number of daily measurements needed to estimate habitual step count levels using wrist-worn trackers and smartphones in 212,048 adults. Sci Rep 2021;11(1):9633 [FREE Full text] [doi: 10.1038/s41598-021-89141-3] [Medline: 33953288]
13. Baranowski T, Mâsse LC, Ragan B, Welk G. How many days was that? We're still not sure, but we're asking the question better!. Med Sci Sports Exerc 2008;40(7 Suppl):S544-S549 [FREE Full text] [doi: 10.1249/MSS.0b013e31817c6651] [Medline: 18562972]
14. Sainani K. Reliability statistics. PM R 2017;9(6):622-628. [doi: 10.1016/j.pmrj.2017.05.001] [Medline: 28602174]
15. Arnardottir ES, Islind AS, Óskarsdóttir M, Ólafsdóttir KA, August E, Jónasdóttir L, Sleep Revolution. The sleep revolution project: the concept and objectives. J Sleep Res 2022;31(4):e13630. [doi: 10.1111/jsr.13630] [Medline: 35770626]
16. Fridgeirsdottir KY, Ólafsdóttir KA, Islind AS, Leppänen T, Arnardottir ES, Saavedra JM. The role of physical activity on obstructive sleep apnea severity and hypoxic load, and the mismatch between subjective and objective physical activity assessments. J Sleep Res 2024;33(6):e14195. [doi: 10.1111/jsr.14195] [Medline: 38480993]
17. Fridgeirsdottir KY, Murphy CJ, Islind AS, Árnadóttir BS, Hrubos-Strøm H, Arnardottir ES, et al. Effects of exercise and a lifestyle app on sleep disordered breathing, physical health and quality of life. ERJ Open Res 2025;11(3). [doi: 10.1183/23120541.01134-2024] [Medline: 40551796]
18. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
19. 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;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Biedebach L, Óskarsdóttir M, Arnardottir E, Islind A. Two sides of the same pillow: unfolding the relationship between objective and subjective sleep quality with unsupervised learning. 2023 Presented at: Forty-Fourth International Conference on Information Systems; 2023 December 10-13; Hyderabad, India URL: https://aisel.aisnet.org/icis2023/ishealthcare/ishealthcare/20/?utm_source=chatgpt.com
21. Atkin AJ, Gorely T, Clemes SA, Yates T, Edwardson C, Brage S, et al. Methods of measurement in epidemiology: sedentary behaviour. Int J Epidemiol 2012;41(5):1460-1471. [doi: 10.1093/ije/dys118] [Medline: 23045206]

22. Altman N, Krzywinski M. Clustering. Nat Methods 2017;14(6):545-546. [doi: 10.1038/nmeth.4299]
23. Prescott S, Traynor JP, Shilliday I, Zanotto T, Rush R, Mercer TH. Minimum accelerometer wear-time for reliable estimates of physical activity and sedentary behaviour of people receiving haemodialysis. BMC Nephrol 2020;21(1):230 [FREE Full text] [doi: 10.1186/s12882-020-01877-8] [Medline: 32546225]
24. Aadland E, Ylvisåker E. Reliability of objectively measured sedentary time and physical activity in adults. PLoS One 2015;10(7):e0133296 [FREE Full text] [doi: 10.1371/journal.pone.0133296] [Medline: 26192184]
25. Gretebeck RJ, Montoye HJ. Variability of some objective measures of physical activity. Med Sci Sports Exerc 1992;24(10):1167-1172. [Medline: 1435166]
26. O'Brien CM, Kitas GD, Rayner F, Isaacs JD, Baker KF, Pratt AG, BIOFLARE Consortium. Number of days required to measure sedentary time and physical activity using accelerometery in rheumatoid arthritis: a reliability study. Rheumatol Int 2023;43(8):1459-1465 [FREE Full text] [doi: 10.1007/s00296-023-05342-1] [Medline: 37227468]
27. Kocherginsky M, Huisingh-Scheetz M, Dale W, Lauderdale DS, Waite L. Measuring physical activity with hip accelerometry among U.S. older adults: how many days are enough? PLoS One 2017;12(1):e0170082 [FREE Full text] [doi: 10.1371/journal.pone.0170082] [Medline: 28081249]
28. Matthews CE, Ainsworth BE, Thompson RW, Bassett DR. Sources of variance in daily physical activity levels as measured by an accelerometer. Med Sci Sports Exerc 2002;34(8):1376-1381. [doi: 10.1097/00005768-200208000-00021] [Medline: 12165695]
29. Tworoger S, Davis S, Vitiello M, Lentz M, McTiernan A. Factors associated with objective (actigraphic) and subjective sleep quality in young adult women. J Psychosom Res 2005;59(1):11-19. [doi: 10.1016/j.jpsychores.2005.03.008] [Medline: 16126091]
30. Knutson KL, Rathouz PJ, Yan LL, Liu K, Lauderdale DS. Intra-individual daily and yearly variability in actigraphically recorded sleep measures: the CARDIA study. Sleep 2007;30(6):793-796 [FREE Full text] [doi: 10.1093/sleep/30.6.793] [Medline: 17580601]
31. Aili K, Åström-Paulsson S, Stoetzer U, Svartengren M, Hillert L. Reliability of actigraphy and subjective sleep measurements in adults: the design of sleep assessments. J Clin Sleep Med 2017;13(1):39-47 [FREE Full text] [doi: 10.5664/jcsm.6384] [Medline: 27707448]
32. Dwyer T, Ponsonby A, Ukoumunne OC, Pezic A, Venn A, Dunstan D, et al. Association of change in daily step count over five years with insulin sensitivity and adiposity: population based cohort study. BMJ 2011;342:c7249. [doi: 10.1136/bmj.c7249] [Medline: 21233153]

Abbreviations

AHI: apnea-hypopnea index
ICC: intraclass correlation coefficient
OSA: obstructive sleep apnea
REDCap: Research Electronic Data Capture

Edited by L Buis; submitted 15.Jul.2024; peer-reviewed by N Maldaner, C Sakal; comments to author 11.Apr.2025; revised version received 15.Jul.2025; accepted 06.Oct.2025; published 10.Feb.2026.

Please cite as:
Murphy CJ, Jouan GM, Friðgeirsdóttir KY, Islind AS, Saavedra JM, Óskarsdóttir M, Arnardóttir ES
Determining Cluster-Specific Differences in the Number of Days Required to Reliably Predict Habitual Physical Activity: Intraclass Correlation Resampling Analysis
JMIR Mhealth Uhealth 2026;14:e64323
URL: https://mhealth.jmir.org/2026/1/e64323
doi:10.2196/64323
PMID:

Designing a Self-Guided Digital Intervention for Self-Management of Shoulder Pain in People Living With Spinal Cord Injury: Tutorial on Using a Person-Based Approach

Verna Stavric[1*], PhD; Nicola L Saywell[1*], PhD; Nicola M Kayes[2*], PhD

[1]Department of Physiotherapy, School of Allied Health, Auckland University of Technology, Private Bag 92006, Auckland, New Zealand

[2]Centre for Person Centred Research, Auckland University of Technology, Auckland, New Zealand

[*]all authors contributed equally

Corresponding Author:
Verna Stavric, PhD
Department of Physiotherapy, School of Allied Health, Auckland University of Technology, Private Bag 92006, Auckland, New Zealand

Abstract

Shoulder pain is prevalent in people living with spinal cord injury. Technology and digital rehabilitation tools are increasingly available, but this has not yet included the provision of a self-guided exercise intervention focused on managing shoulder pain for people living with spinal cord injury. We drew on the person-based approach (PBA) to intervention development to design a Shoulder Pain Intervention delivered over the interNet (SPIN) to address this gap. However, in preparation for the design process, we found very few published examples of how the PBA had been operationalized. The aim of this paper is to provide a detailed explanation of our approach and how we operationalized the PBA in the design of SPIN to maximize relevance and engagement. Our design process followed the key PBA steps, combining additional evidence and theoretical components. Each step ensured that guiding principles were formulated and followed to maximize the probability that SPIN would be fit for purpose. We followed 3 steps: (1) we drew on themes from preparatory research (existing and primary) to identify the key behavioral issues, needs and challenges, and existing features to form the basis of SPIN design; (2) we formatted guiding principles that included articulating specific design objectives to provide a framework to identify system requirements; and (3) we selected and refined intervention features using existing literature, behavioral theory, and tools such as the "Behaviour Change Wheel." We have designed SPIN by incorporating a deep understanding of the users' needs and best available evidence to maximize engagement and positive outcomes. In this paper, we have made clear how we operationalized the PBA phases, including how existing evidence, theory, tools, and methods were leveraged to support the PBA process. In explicating our process, we have provided a blueprint to guide future researchers using this approach.

(JMIR Mhealth Uhealth 2026;14:e66678) doi:10.2196/66678

KEYWORDS

person-based approach; self-guided; intervention design; behavioral analysis; spinal cord injury; shoulder pain; self-management; mHealth; mobile apps; smartphones; digital health; digital interventions

Background

Overview

Shoulder pain is common in wheelchair users living with spinal cord injury (SCI) [1,2]. A lesion to the spinal cord can result in loss of innervation to muscles of the trunk and lower limbs. Consequently, many people living with SCI (pwSCI) rely on their upper extremities not only for performance of daily activities but also for locomotion. Shoulder pain can have a significant impact on their activity, reducing mobility, independence, and quality of life [1-5]. Digital and web-based interventions have increasingly been offered to pwSCI to promote exercise and physical activity [6-9]. These interventions minimize barriers to rehabilitation to address many health concerns, including managing their shoulder pain. Previous authors have found that in the general population, digital or web-based interventions can produce positive effects in various outcomes, such as physical activity [10,11].

Technology-supported exercise interventions for pwSCI with persistent shoulder pain are currently available, but they have some limitations. They either require ongoing input and monitoring from a clinician [12-15] or provide general self-management advice [13] but without enough guidance to allow for clear and structured exercise progression specifically for shoulder pain. Self-guided digital exercise interventions have been successfully implemented for people with knee osteoarthritis [16,17], dizziness [18,19], and breast cancer [20] and may be a viable option for pwSCI. Our recent systematic

review and meta-analysis of self-guided digital physical activity and exercise interventions demonstrated positive effects on physical activity at both short- and longer-term follow-up, in people living with chronic conditions [21]. We also found that interventions that used behavioral strategies and were underpinned by a theoretical framework were more effective. This suggests that self-guided digital interventions have the potential to support pwSCI to manage their shoulder pain, but that the intervention would need to be designed systematically and intentionally.

We have designed Shoulder Pain Intervention delivered over the interNet (SPIN) as a self-guided digital intervention to give pwSCI who experience shoulder pain the ability to access and progress evidence-based exercises. The intervention guides pwSCI to monitor symptoms and improvement [22] to promote autonomy in the management of their condition. The aim of SPIN is to be an engaging program that is responsive to the needs of pwSCI who have shoulder pain.

To achieve this, we were guided by the person-based approach (PBA) in the design of SPIN [23]. The PBA follows 4 iterative phases of intervention development that include (1) planning which seeks a deep understanding of the perspectives and psychosocial context of potential users through iterative qualitative research, (2) design based on guiding principles that have been created from insights from the first phase, (3) development and refinements which are made through iterative user feedback, and (4) trialing to evaluate the effectiveness on outcomes and impact on behavior change to make any necessary adjustments. Due to its focus on the development of digital behavior change interventions, the intent and purpose of PBA align well with adjacent behavior change theory and tools such as the COM-B [24], "Behaviour Change Wheel" [25], and behavioral analysis [26]. Furthermore, the PBA process is sufficiently flexible to enable the use of these (and other) tools to achieve the aims and purpose of a given phase. Integrating behavioral science theory and evidence while keeping users' needs and contexts in focus has been found to maximize engagement and effectiveness of interventions [18,25,27-31]. This tutorial focuses on the first 2 PBA phases of planning and design. See Table 1 for an example of how our study was mapped onto the PBA.

Table . Mapping of person-based approach phases onto Shoulder Pain Intervention delivered over the Internet design.

PBA[a] description	Phase	This study	
		Purpose	Planned outcome
Use of primary and secondary *qualitative evidence* to understand users' behavioral and psychosocial needs and challenges in using the intervention		To determine *factors that need to be included* to *encourage or facilitate engagement* with this self-guided web-based exercise intervention	• A rich description of key *needs, challenges, and facilitators* of engagement in web-based tools and exercise for people living with SCI[b] who experience shoulder pain to *underpin the design phase's guiding principles and features*
Formulation of *key guiding principles* that capture the main intervention objectives as identified in the planning phase and that are continuously referred to throughout the development of the intervention		To design an *evidence-based, self-guided, web-based* intervention Exercise, behavioral support, and self-guided components to be included within the intervention features	• Intervention design objectives • Intervention features • First iteration of SPIN[c] prototype

[a]PBA: person-based approach.

[b]SCI: spinal cord injury.

[c]SPIN: Shoulder Pain Intervention delivered over the interNet.

We drew heavily on the existing PBA literature during the planning and designing stages of SPIN. However, the lack of access to detailed examples of how the PBA has been operationalized in practice made it challenging to translate the principles of this approach into reality. This is not a unique problem. Duncan and colleagues [32] noted that published work on the development of an intervention is frequently sparse because it is often included in the same publication as the reporting of a pilot or feasibility study.

Aim

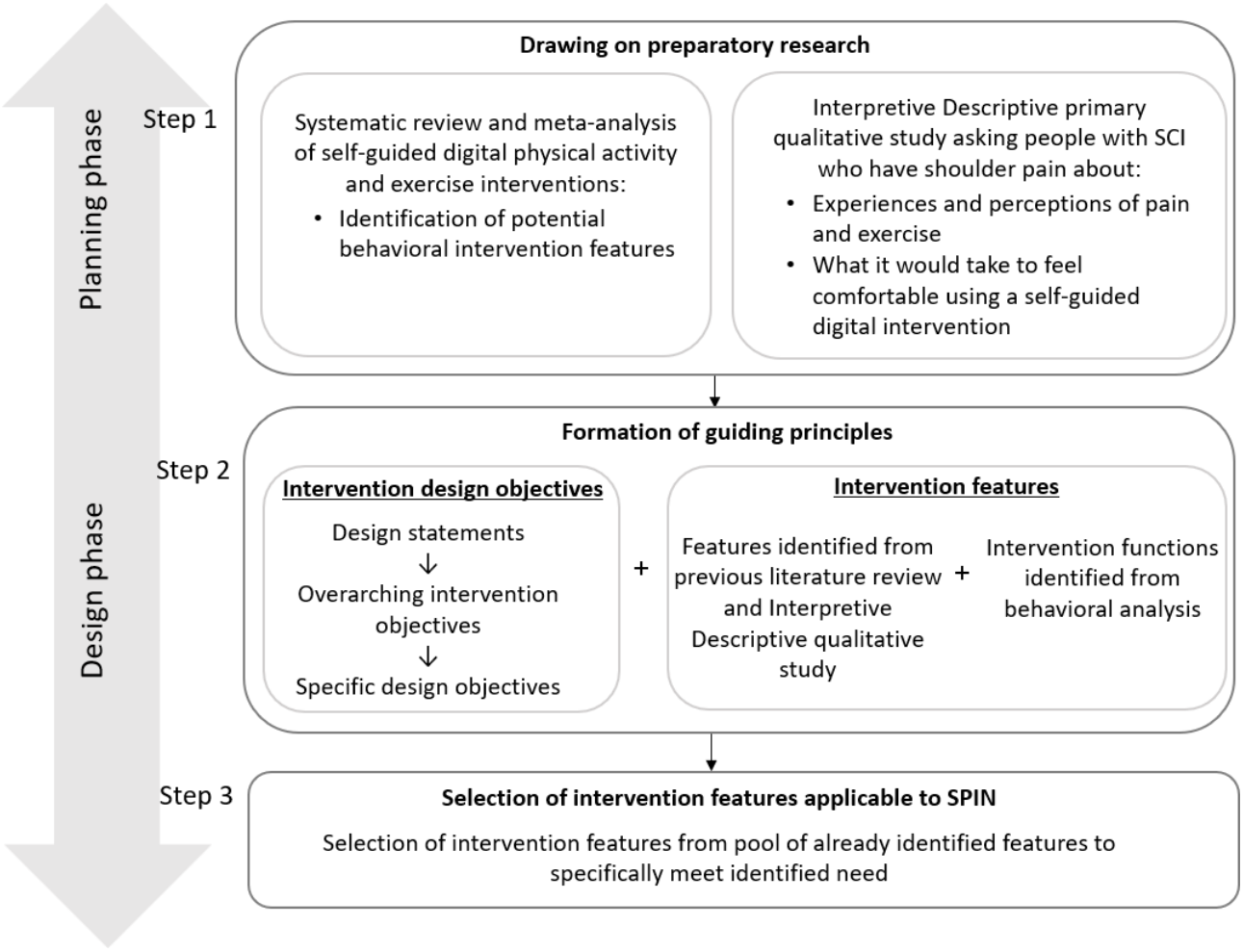
The aim of this paper is to make clear how the principles of the PBA were operationalized in intervention design and the development of SPIN. We have illustrated our use of the PBA framework by outlining the detailed and explicit steps involved in the translation of the evidence, theory, and person-based recommendations into intervention design. In doing so, we have built on the existing methodological framework and enabled others to draw on this approach in future intervention design and development.

Methods and Outcomes

Overview

The planning and design phases of the PBA are described below, along with an overview of how they were operationalized in the design of SPIN. Figure 1 provides an overview of the SPIN design process and the components involved. Each step and its subsequent outcome have been described in detail in the sections that follow.

Figure 1. Overview of Shoulder Pain Intervention delivered over the Internet design steps and components. SCI: spinal cord injury; SPIN: Shoulder Pain Intervention delivered over the interNet.



Step 1: Drawing on Preparatory Research

Methods

This initial phase of the PBA draws on qualitative research, including interviews and focus groups, to gather in-depth insights into the psychological, social, and emotional factors that influence the users' behavior. The goal is to identify the underlying motivations, beliefs, and barriers that may affect engagement with health interventions [23]. In the context of SPIN, this preparatory research included: (1) a systematic review and meta-analysis investigating the effectiveness of self-guided digital physical activity and exercise interventions [21] and (2) an Interpretive Descriptive qualitative study exploring the perceptions of pwSCI who have shoulder pain, on the use of a self-guided digital intervention to help them manage their shoulder pain [33].

Outcome

The review identified several self-guided digital physical activity and exercise interventions. Data extraction included identifying discrete intervention features and categorizing them using a purpose-built template (Multimedia Appendix 1), based on a synthesis of key literature [27,34-40]. Using this template, we extracted possible behavioral intervention features relating to qualities such as customizability, the provision of instruction, feedback and monitoring, tailoring, reminders and prompts, goals and planning, social support, and rewards and threats. We also noted the success of interventions using features that supported behavior, particularly self-regulation. This informed an initial pool of possible intervention features for SPIN that were reviewed later in Step 2.

The Interpretive Descriptive qualitative study identified themes that represented an evaluative process pwSCI go through when considering using a self-guided digital exercise intervention: *Should I use it?*, whether I believe it will work for me right now; *Can I use it?*, whether I can operate the intervention competently and confidently; and *Will I use it?*, whether it will be responsive to my unique needs and keep me coming back. These formed the basis of the design statements in Step 2.

Conceptual representations of possible behavioral intervention features identified from the review were used as probes and images during data collection in the Interpretive Descriptive qualitative study. These were used to prompt discussion about what could help pwSCI to engage in a self-guided digital intervention. The pwSCI discussed ways in which these concepts and specific features may support them. These perspectives were extracted from the audio recordings and tabulated to support the identification of behavioral intervention features in Step 2.

Step 2: Formation of Guiding Principles

Overview

The guiding principles in the PBA are formulated by synthesizing key insights from the planning phase (Step 1) into intervention design objectives and corresponding intervention features that address users' specific needs, preferences, and behavioral barriers [23]. Yardley and colleagues [23] contend that staying true to the identified needs of the people who will use the intervention, throughout the design process, increases intervention relevance, engagement, and effectiveness. In our design of SPIN, we followed several stages to ensure the key context-specific behavioral needs and challenges identified in the Interpretive Descriptive qualitative study remained the focal point during intervention design.

Intervention Design Objectives

Yardley et al [23] suggest generating intervention design objectives to support the creation of the guiding principles but do not expand on how these may be identified. Below, we describe the method we followed to produce intervention design objectives through the creation of design statements, overarching intervention objectives, and specific intervention design objectives.

Design Statements

Methods

We created design statements by using the 3 themes constructed in the Interpretive Descriptive qualitative study. We first reframed each theme into a design statement, giving consideration to how each could be reflected in the design of the intervention. To do this, we reworded the themes to move from a question (*Should I use it?*) into a design statement (*I should use it if…*) and then added conditions applicable to each design statement. Each condition reflected key elements from the qualitative findings, resulting in person-centered conditions to be met in the design process. This process provided depth and context to inform the design of SPIN and ensured the next step would be underpinned by the perspectives of the future users of the intervention, in this case, pwSCI.

Outcome

The Interpretive Descriptive qualitative study themes, design statements, and key conditions for success are presented in Table 2.

Table . Translation of themes to design statements and conditions of success.

Interpretive Descriptive qualitative study theme	Reframed to: design statements	Conditions for success
Should I use it?	I should use it if:	• I believe it will work for me • There is evidence of credibility • There is a clear indication that it is suitable for me • It resonates with my current attitude toward exercise, support situation
Can I use it?	I can use it if:	• I can use it competently • I can use it confidently • It can be tailored and adapted to my unique needs • I can use it safely, without causing more harm • I have the belief that I could use it, given the resources and capacity I have • I have the right support to use it
Will I use it?	I will use it if:	• It is responsive to my unique needs • It encourages me to progress when I am ready • I feel supported to use it • I can see progress as a consequence of using it • It keeps me coming back

Overarching Intervention Objectives

Methods

Next, we articulated the overarching intervention objectives. Succinctly describing the intervention objectives allows a snapshot of the key characteristics of the intervention [23]. We, therefore, clearly articulated how SPIN is distinctive and different from other interventions, reflecting the specific behavioral issues, needs, and challenges it must address.

We developed the intervention objectives iteratively, repeatedly revising the wording with reference to the original research question and design statements, and with input from the research team and stakeholders. Stakeholders included pwSCI, a clinician with experience in SCI rehabilitation, a clinician who was also a pwSCI and a representative of a relevant nongovernmental organization. Each iteration strived to reflect the essence of the needs expressed by the participants with wording that represented what ideal uptake and use of this self-guided digital exercise intervention could look like. The overarching

intervention objectives were then used as a reference point for later design and development phases.

Outcome

Referring to the design statements and overall research aim, the overarching intervention objectives for SPIN were to:

1. Be tailored to users' specific and unique needs so they can relate to it and trust it and so that it can be responsive to their changing needs while using SPIN; and
2. Enable users to use it competently and confidently within their capabilities and support systems in a way that is safe and motivating.

Specific Design Objectives

Methods

Once the overarching intervention objectives were formulated, we created the specific design objectives underpinned by the design statements. We developed a working definition, incorporating the key conditions for success for each specific design objective, to ensure clarity in interpretation. These were then reviewed against the overarching intervention objectives, making sure they supported the overall objectives of SPIN. We continued to refine them as the design process progressed, during our planned discussion forums.

Outcome

Tables 3-5 each refer to a different theme. Specific design objectives and working definitions are presented in the first 2 columns; intervention functions and features are discussed in later sections.

Table . Guiding principles from the theme *Should I use it?*.

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To help users relate to and trust the program	The program will give users confidence in the source, message, and value of the program. The program is credible and legitimate and promotes trust.	• Education • Training • Modeling • Enablement • Persuasion	• Development team details (names, credentials, and contact info) • Endorsements • *Testimonials* (source matching for social comparison) • Evidence for shoulder pain exercises • How user data will be used or stored • Professional polished interface and function
To reassure users it will be clear who the program is suitable for, giving users confidence that the program is right for them and at what stage it is right for them	The program will guide users through a process to be able to screen for and identify if they are suitable to use the intervention and to promote trust and confidence that this is a safe and robust process.	• Education • Training • Modeling • Enablement • Persuasion	• Screening questionnaire/questions (that will exclude those unsuitable) • Monitoring questions at each exercise event and tracking this information • *FAQ*[a] *section* • Contact information for the team
To provide a sense of potential that it will work for them	The program will help users identify with it and the potential that it may have for them, in their current situation.	• Education • Training • Modeling • Enablement • Persuasion	• *Testimonials* (image with text, video, and quotes) of people in different "stages" of readiness or different situations. • FAQ section addressing suitability of different situations "Is this right for me?" or "How do I know this is right for me?" or "Questions I can ask to make sure this is right for me?"

[a]FAQ: frequently asked question.

Table . Guiding principles from theme *Can I use it?*.

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To promote a sense of safety when using the program	The program will ensure exercises are at the appropriate difficulty level and will be responsive to changes in user presentation to ensure that they don't significantly aggravate shoulder symptoms.	• Training • Environmental restructuring • Modeling • Enablement	• *Monitoring and tracking of shoulder pain* and exercise difficulty • Exercise selection based on user responses and a priori rules • Program-generated advice based on user responses, such as acknowledging concerns, referral to *FAQ*[a], evidence, health care provider
To promote user competence	The program will be easy to use by a range of users and in a range of circumstances, giving them a sense of confidence when using it in the context of their unique life situation.	• Training • Environmental restructuring • Modeling • Enablement	• Language at an appropriate reading level • Layout is clear and simple • Font size and buttons are large for reduced hand function • Minimal scrolling and clicking • Consistent screen layout • Clear signposts • Logical interface • Exercises presented in video and audio formats by pwSCI • Exercises presented in step-by-step processes • Exercises are planned to fit in with daily routine and normal digital device use • Tunneling of information (releasing information in small amounts, as the user progresses through "right amount, at the right time") • Graded goal setting, implementation planning • *Tailored* and action feedback based on tracking • Praise for success • Advice or support if not yet succeeded • Digital use guidance when needed (help link)
To promote user autonomy	The program will give users a sense of control and ownership over the program and their progress through the program.	• Training • Environmental restructuring • Modeling • Enablement	• *Offering choice where possible: tailoring functions* in exposure matching-timing, intensity (when and how often) • *Reminders* • Excercise selection, timing of exercise • Intervention delivery • Tunneling of options into the most common choices • Suggestions or options for different situations

[a]FAQ: frequently asked question.

Table . Guiding principles from the theme *Will I use it?*.

Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To promote a positive emotional experience	The program will incorporate positive autonomy-supportive language that invites, informs, and supports users to work through the program.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Use of *positive language and tone* in inviting users to decide for themselves "some find it helpful." • Use of anecdotes to describe examples of success, decision-making • Acknowledging and addressing concerns about using the program, such as pain or carer support • Using *FAQ*[a] *section* • Use of useful/interesting/relevant/personal reminders • Positive or encouraging wording on feedback on progress toward the goal
To promote a sense of relatedness	The program will be relevant to the user by using communication and wording that is tailored to their self-identified preferences and personalized to their unique circumstances.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Feedback as above (and that is immediately reciprocated when interacting with the intervention) • Competition with others, and/or • Cooperation with others • Social connection through the program's grouping • Initial "getting to know you" questionnaire to help with *tailoring ingredients* • Personalization: (1) identification (including username in correspondence), (2) raising expectation (including relevant information in correspondence that is based on users' responses to questions/input), and (3) contextualization (*wording, examples that are relevant to user*-exercises relevant for tetra vs para) • *Reminders* • *Testimonials* • Self-identified support
To help users maintain their exercise over the 12 weeks	The program will use a variety of strategies and features to encourage and support users to maintain engagement in their exercise for the duration of the program.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Rewards (points or similar)/competition • Goal setting • Action planning • Communication that is *positive*, immediate, and useful and *tailored*

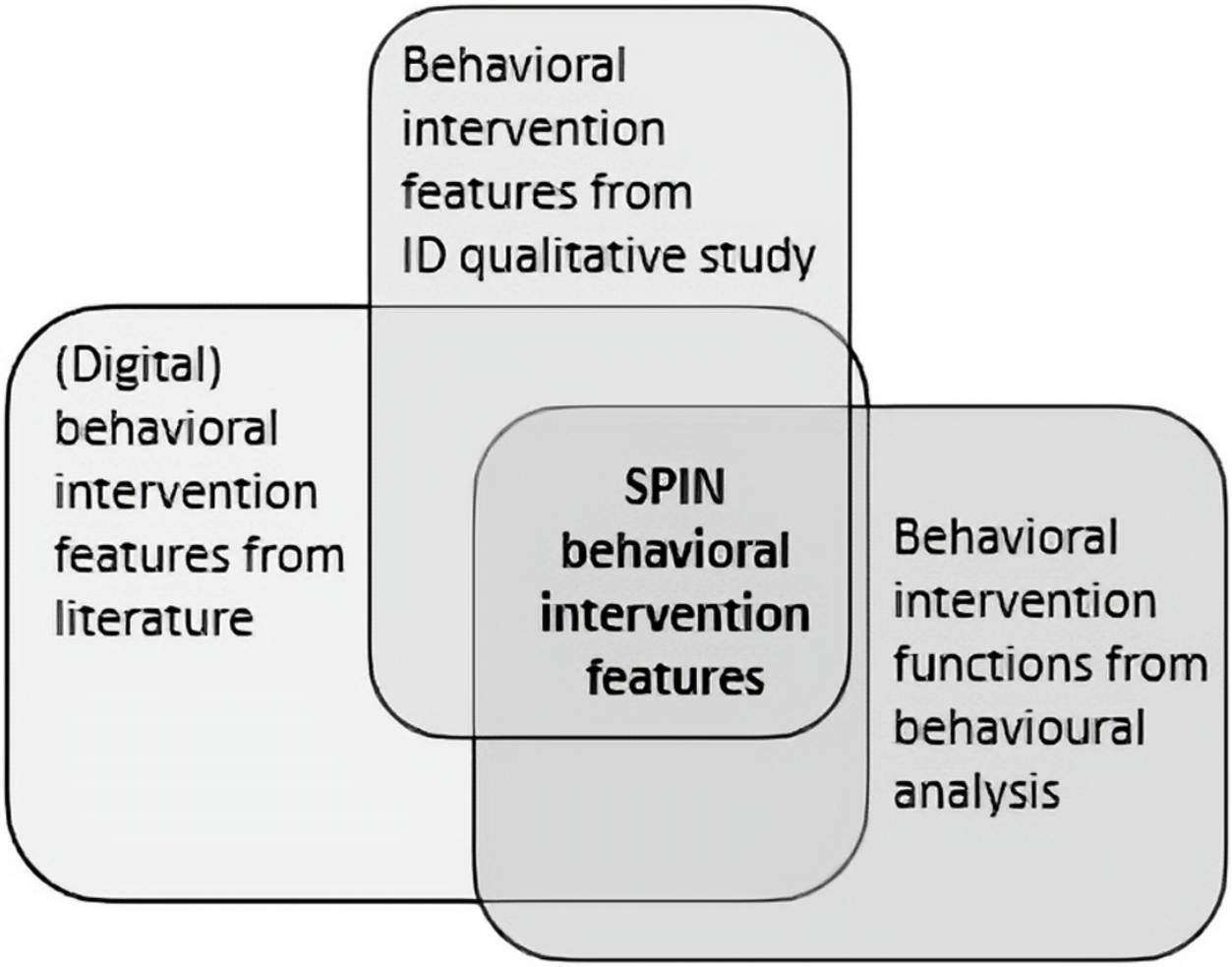
Design objectives that address identified needs, issues, and challenges	Working definition	Intervention functions	Intervention features that address the design objectives
To promote a sense of accountability	The program will provide features that encourage the user to return to the program and to continue with the exercises.	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Competition with others or with self • Support from others • Communication that is *positive*, immediate, and useful and *tailored* • Communication that is personalized • Rewards that are only released upon completion of a certain amount of exercise
To promote a sense of progree and engagement	The program will enable the user to understand their progress through a clear and simple tracking feature. This will be done in a way that encourages further progress and ongoing engagement with the exercise intervention	• Training • Environmental restructuring • Enablement • Modeling • Education • Persuasion • Incentivization	• Feedback and tracking • Choice in exercise selection • Personalization • *Tailoring*

[a]FAQ: frequently asked question.

Intervention Features

In the PBA, the guiding principles inform the intervention features by providing a framework for selecting and shaping features that directly support the specific design objectives, and to improve resonance, engagement, and acceptability of an intervention [23]. A range of evidence informed the selection of behavioral intervention features: (1) in our review, we identified a range of features used in digital interventions that have been associated with better health-related outcomes [27,34-36]; (2) we identified possible behavioral intervention design features from our Interpretive Descriptive qualitative study [33]; and (3) we identified behavioral "intervention functions" we were trying to achieve using a behavioral analysis as per Michie and colleagues' framework [26]. We then mapped these to the most relevant intervention features. Figure 2 represents the layers of evidence that informed SPIN's intervention features. We will describe each of these in detail below.

Figure 2. Layers of evidence that informed Shoulder Pain Intervention delivered over the Interjet's intervention features. ID: Interpretive Descriptive; SPIN: Shoulder Pain Intervention delivered over the interNet.



Identifying Behavioral Intervention Features From Previous Literature Review and the Interpretive Descriptive Qualitative Study

Methods

In Step 1, we had earlier identified potential behavioral intervention features for self-guided interventions that were identified from our systematic review and meta-analysis, using the specifically developed template, drawing from the CONSORT-EHEALTH checklist (V.1.6.1) [38]. See Multimedia Appendix 1 for a sample of our template showing sections used to record behavioral intervention features. For this current stage of the SPIN design, we also reviewed intervention features of publications that missed the strict inclusion criteria of the systematic review and meta-analysis but addressed digital delivery of physical activity or exercise intervention for possible relevant behavioral intervention features. We then integrated the data on specific features collected from our Interpretive Descriptive qualitative study. These data were categorized by proposed purpose and function and then mapped against the specific design objectives.

Outcome

There was overlap, resulting in some features identified as addressing more than one design objective. Many of the studies in the systematic review included digital behavioral intervention features that involved instruction on exercise or physical activity performance, self-monitoring of the exercise or physical activity behavior, goals and planning, and prompting. The results of the Interpretive Descriptive qualitative study and other reviewed literature suggested additional behavioral intervention features. Table 6 presents a summary of the behavioral intervention feature categories that we considered for SPIN, the design objective(s) they are related to, and the supporting evidence.

Table . Behavioral intervention feature categories supported by systematic review, Interpretive Descriptive qualitative study, and existing literature.

Design objectives	Behavioral intervention feature	Proportion of studies identified in systematic review and meta-analysis (out of 16 studies)	Identified in Interpretive Descriptive qualitative study	Identified in other literature not included in meta-analysis
Should I use it?	Ensuring personal relevance	16	✓	Horsch et al [41]
Should I use it?/Can I use it?	Use of credibility and trust-enhancing features	5	✓	Bossen et al [42]; Oinas-Kukkonen and Harjumaa [43]
Should I use it?/Will I use it?	Provision of information about actual users	2	✓	Morrison et al [34]
Can I use it?	Allowance of the user to control or adapt features	7	✓	McClure et al [44]
Can I use it?	Ensuring ease of use	6	✓	Carter et al [45]; Hurling et al [46]; Webb et al [27]
Can I use it?	Provision of information 'just in time' and in 'just the right amount'	9	✓	Oinas-Kukkonen and Harjumaa [43]; Xu et al [47]
Can I use it?/Will I use it?	Use of goal setting	8	✓	Webb et al [27]; Willett et al [48]; Dugas et al [49]
Can I use it?	Use of demonstration of behavior	10	✓	Webb et al [27]
Can I use it?	Use of feedback of behavior	10	✓	Webb et al [27]; Dugas et al [49]
Can I use it?	Use of tailored feedback	10	✓	Morrison et al [35]; Dugas et al [49]
All	Use of tailoring based on a number of variables	5	✓	Morrison et al [34]; Couper et al [50]; Xu et al [47]; Oinas-Kukkonen and Harjumaa [43]; Figueiras and Neto [51]; Dugas et al [49]
Can I use it?/Will I use it?	Use of reminders	8	✓	Webb et al [27]; Lin and Wu [52]; Alahäivälä and Oinas-Kukkonen [53]; Dugas et al [49]
Can I use it?/Will I use it?	Use of self-monitoring features	9	✓	Morrison et al [34]; Glasgow et al [54]; Willett et al [48]
Can I use it? /Will I use it?	Use of positive tone and language	4	✓	Haines-Saah et al [55]
Can I use it? /Will I use it?	Use of text message	3	✓	Webb et al [27]
Will I use it?	Use of action/coping planning	5	✓	Webb et al [27]; Glasgow et al [54]; van Genugten et al [56]
Will I use it?	Use of facilitation of social comparison and support	2	✓	Webb et al [27]; Davies et al [57]; Perski et al [58]; Alahäivälä and Oinas-Kukkonen [53]; Xu et al [47]
Will I use it?	Use of rewards and incentives	1	✓	Khadjesari et al [59]; Schubart et al [60]; van Genugten et al [56]
All	Use of a combination and a number of features	14		Webb et al [27]; Meade et al [61]

Identifying Intervention Functions From a Behavioral Analysis

Methods

We included a behavioral analysis using the "Behaviour Change Wheel" and COM-B model as outlined by Michie and colleagues [26]. This is a theoretical framework that provides a systematic way of identifying the problem and analyzing the behavioral needs of a target behavior. The "Behaviour Change Wheel" can support intervention design by linking the identified behavioral needs to "intervention functions" through a mechanism of action.

Consistent with the guiding principles and specific design objectives, and for the purpose of this behavioral analysis, we reframed the 3 themes from the Interpretive Descriptive qualitative study into target behaviors: *Should I use it?*—Signing up to SPIN (Table 7); *Can I use it?*—Using SPIN (Table 8); and *Will I use it*?—Returning to SPIN over the 12 weeks (Table 9). The COM-B Model was then used to identify the capability (C), opportunity (O), and motivational (M) components required for each of these behaviors (B) to occur, referring to the specific design objectives. The questions "*what needs to happen for the target behavior to occur?"* and "*is there a need to change?"* facilitated the analysis process [26]. We used this process to identify (or "diagnose") the relevant COM-B components that need to be addressed for the target behavior to occur (see the Behavioral diagnosis of the relevant COM-B components in Tables 7-9).

Table . Behavioral analysis of target behavior: signing up to SPIN[a] (*Should I use it?*) for people living with spinal cord injury who have shoulder pain.

COM-B components[b]	What needs to happen for the target behavior to occur?	Is there a need for change?
Physical capability	Have the physical ability to access SPIN features and functions and use it	No change needed as SPIN will only be suitable for people who can physically access and use it
Psychological capability	Believe they have the capability to use SPIN	*Change* needed as pwSCI[c] will want reassurance that they have sufficient physical capability to use SPIN and/or that it is suitable for people with their level of physical ability
Psychological capability	Know that exercise can improve pain symptoms (or not make the condition worse)	*Change* may be needed as there may be fears or concerns that exercise could worsen pain symptoms
Physical opportunity	Have a device that can access SPIN	No change needed as SPIN will only be suitable for those people who have devices that can access SPIN
Social opportunity	Know about other pwSCI who have either benefitted from exercise for shoulder pain or are using SPIN	*Change* needed as pwSCI may not know about others who have benefitted from exercise to improve shoulder pain symptoms or who are using SPIN
Reflective motivation	Hold beliefs that exercising will reduce pain symptoms and/or improve activity	*Change* needed as pwSCI may be fearful that exercise may worsen pain symptoms
Reflective motivation	Believe that SPIN has been developed by a credible and trustworthy source	*Change* needed as pwSCI will want to assure themselves that SPIN has been developed by knowledgeable personnel who have experience in SCI[d] rehabilitation
Automatic motivation	Believe that SPIN will identify those that are suitable (and unsuitable) to use it	*Change* needed as pwSCI will want assurance that SPIN is appropriate for their circumstances and can be tailored for their needs
Automatic motivation	Need to feel that SPIN resonates (with current attitude toward exercise, support situation)	*Change* needed as pwSCI need to feel comfortable that SPIN is right for them at this time
Behavioral diagnosis of the relevant COM-B components	Psychological capability, social opportunity, reflective and automatic motivation need to change for the target behavior to occur	—[e]
Likely ”intervention functions” that link to COM-B	Education (psychological capability, reflective motivation), Training (physical opportunity), Modelling (social opportunity), and Persuasion (reflective motivation, automatic motivation)	—

[a]SPIN: Shoulder Pain Intervention delivered over the interNet.

[b]Behavioral diagnosis of the relevant COM-B components: psychological capability, social opportunity, reflective and automatic motivation need to change for the target behavior to occur.

[c]pwSCI: people living with spinal cord injury.

[d]SCI: spinal cord injury.

[e]not applicable.

Table . Behavioral analysis of target behavior: using SPIN[a] (*Can I use it?*) for people living with spinal cord injury who have shoulder pain.

COM-B components[b]	What needs to happen for the target behavior to occur?	Is there a need for change?
Physical capability	Have the physical ability to control and manipulate SPIN features and functions and related equipment and setup	*Change* may be needed as pwSCI[c] will want reassurance that they have sufficient physical capability to use the intervention and/or that the intervention is suitable for people with their level of physical ability
Physical capability	Have the additional support as required	*Change* may be needed with additional support for equipment setup and exercise support
Psychological capability	Believe they have the capability to use SPIN	*Change* needed as pwSCI will want reassurance that they have sufficient physical capability to use SPIN and/or that it is suitable for people with their level of physical ability
Psychological capability	Know how to navigate through the intervention	*Change* needed to clearly provide pwSCI with signposts and information to guide them through
Psychological capability	Know how to perform exercises safely	*Change* needed to ensure appropriate level of exercises is offered and explained to maximize safe exercising and to ensure that the program is responsive to changes in user presentation
Physical opportunity	Have a program that is usable and easy to follow	*Change* needed to ensure SPIN is easy to use and understand
Social opportunity	Haencouragement from peers	*Change* needed to ensure access to a community of users
Reflective motivation	Have confidence in one's ability to use the intervention program	*Change* needed to provide a sense of ownership and control of the program, with positive reinforcement with use
Reflective motivation	Have belief the intervention will enable achievement of outcomes important to user	*Change* needed as users may not recognize the value of SPIN
Automatic motivation	Have experience of benefit from intervention and sense of progress	*Change* needed to provide consistent exercise opportunities
Behavioral diagnosis of the relevant COM-B components	Physical and psychological capability, physical and social opportunity, and reflective motivation need to change for the target behavior to occur	—[d]
Likely "intervention functions" that link to COM-B	Training (physical capability, psychological capability), Environmental restructuring (physical opportunity), Modelling (social opportunity), and Persuasion (reflective motivation)	—

[a]SPIN: Shoulder Pain Intervention delivered over the interNet.

[b]Behavioral diagnosis of the relevant COM-B components: physical and psychological capability, physical and social opportunity, and reflective motivation need to change for the target behavior to occur.

[c]pwSCI: people living with spinal cord injury.

[d]not applicable.

Table . Behavioral analysis of target behavior: using SPIN[a] (*Will I use it?*) for people living with spinal cord injury who have shoulder pain.

COM-B components[b]	What needs to happen for the target behavior to occur?	Is there a need for change?
Physical capability	Have the physical ability to control and manipulate SPIN features and functions and related equipment and setup	*Change* may be needed as pwSCI[c] will want reassurance that they have sufficient physical capability to use the intervention and/or that the intervention is suitable for people with their level of physical ability
Physical capability	Have the additional support as required	*Change* may be needed with additional support for equipment setup and exercise support
Psychological capability	Believe they have the capability to use SPIN	*Change* needed as pwSCI will want reassurance that they have sufficient physical capability to use SPIN and/or that it is suitable for people with their level of physical ability
Psychological capability	Know how to navigate through the intervention	*Change* needed to clearly provide pwSCI with signposts and information to guide them through
Psychological capability	Know how to perform exercises safely	*Change* needed to ensure appropriate level of exercises is offered and explained to maximize safe exercising and to ensure that program is responsive to changes in user presentation
Physical opportunity	Have a program that is usable and easy to follow	*Change* needed to ensure SPIN is easy to use and understand
Social opportunity	Have encouragement from peers	*Change* needed to ensure access to a community of users
Reflective motivation	Have confidence in one's ability to use the intervention program	*Change* needed to provide a sense of ownership and control of the program, with positive reinforcement with use
Reflective motivation	Have belief the intervention will enable achievement of outcomes important to user	*Change* needed as users may not recognize the value of SPIN
Automatic motivation	Have experience of benefit from intervention and sense of progress	*Change* needed to provide consistent exercise opportunities
Behavioral diagnosis of the relevant COM-B components	Physical and psychological capability, physical and social opportunity, and reflective motivation need to change for the target behavior to occur	—[d]
Likely "intervention functions" that link to COM-B	Training (physical capability, psychological capability), Environmental restructuring (physical opportunity), Modelling (social opportunity), and Persuasion (reflective motivation)	—

[a]SPIN: Shoulder Pain Intervention delivered over the interNet.

[b]Behavioral diagnosis of the relevant COM-B components: physical and psychological capability, physical and social opportunity, and reflective motivation need to changefor the target behavior to occur.

[c]pwSCI: people living with spinal cord injury.

[d]not applicable.

Next, we mapped these components to established 'intervention functions,' using the "Behaviour Change Wheel." Most relevant "intervention functions" were then identified from the matrix of links between COM-B and intervention functions [26]. The "Behaviour Change Wheel" uses the term "intervention function" in lieu of intervention "type" or "category" since the same intervention feature may address more than 1 function [26].

Outcome

Tables 7-9 present the target behavior for each design objective and what (if any) change is needed to occur based on the COM-B components. "Intervention functions" most likely to support behavior change have also been identified. For example, testimonials about positive experiences of using exercise to help with shoulder pain could be a form of modeling (providing an example for people to aspire to) and persuasion (using communication to induce positive feelings or stimulate action). This mapping process allowed each specific design objective to be checked to ensure it was supported by an appropriate "intervention function" and corresponding intervention feature. "Intervention functions" linked to the target behavior have been included in Likely "intervention functions" that link to the COM-B in each of the tables (Tables 7-9). The guiding principles tables (Tables 3-5) provide an overview of how these

"intervention functions" map to the design objectives ("Intervention functions" column).

Step 3: Selection of Intervention Features Applicable to SPIN

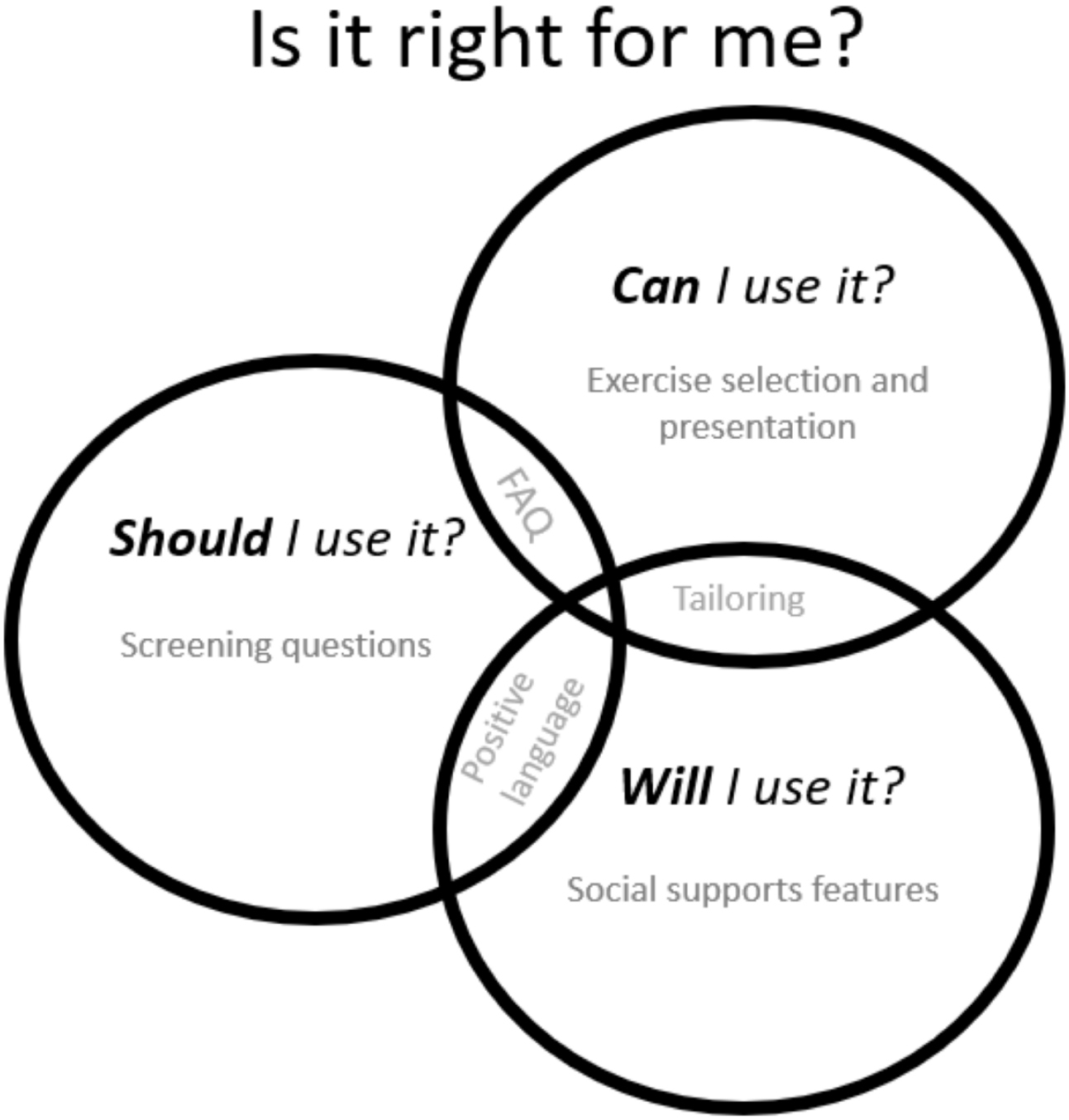
Methods

The design phase of the PBA involves identifying intervention features and content, guided by the previously formulated guiding principles, to ensure alignment with users' psychosocial contexts and to enhance relevance, acceptability, and engagement through iterative user feedback [23]. We were able to begin selecting specific SPIN intervention features once the behavioral analysis was complete. The behavioral intervention features previously identified (Table 6) were reviewed. We mapped those that we felt were contextually appropriate against the "intervention functions." Each was checked to ensure it supported the specific design objectives and the overarching intervention objective. VS completed this process in consultation with coauthors.

Outcome

Collectively, Tables 3-5 demonstrate a complete representation of the guiding principles of SPIN's proposed intervention features and functions, mapped back to the design objectives. Some intervention features address more than 1 intervention design objective. These features have been italicized in Tables 3-5. For example, having a forum for frequently asked questions may reduce barriers to starting the intervention and give users the information they need to progress. Having positive, encouraging language can attract users to start using the intervention and motivate them to continue with it. Other intervention features more clearly support only one of the intervention design objectives. Figure 3 schematically presents an example of how overlapping intervention features cohesively support SPIN's identified design objectives.

Figure 3. Schematic representation of overlapping intervention features.



Application of Our Design Steps to Future Intervention Design

We believe that by explicating how we used the PBA in the development of SPIN, we can support others to use the PBA in the design of interventions. Table 10 provides a summary view of our process and includes some questions that we hope will prompt other researchers to consider how they might operationalize the use of PBA in their work. The table provides an overview of key phases of PBA and possible timelines (column 1) and examples from SPIN (column 2), including tools and methods we drew on as complementary to PBA and which we found useful in operationalizing the approach. In column 3, we have included our reflections on the benefits of our approach. The final column has questions that we hope will serve as prompts for researchers and designers when using this approach.

Table . Operationalizing the person-based approach: our experience and future applications.

1. Key steps in the person-based approach	2. Methods we used to operationalize PBA[a] steps in the development of SPIN[b]	3. Strengths and opportunities of our approach	4. Questions to consider when planning this step
Step 1 (months 3 - 6) Identify key behavioral issues (access), needs (not feeling competent), and challenges the intervention must address	Interpretive Descriptive qualitative study to explore user perspectives of self-guided exercise intervention and what would help or hinder uptake of a self-guided digital exercise intervention. Used probes and images during data collection to help users visualize and provide feedback on possible intervention features.	Drawing on Interpretive Descriptive as a nested study within the PBA process helped to provide a robust framework to capture and make sense of user needs and preferences. Interpretive Descriptive is congruent with the goals of PBA and has the benefit of (1) being oriented toward translation from the outset, (2) prioritizing the production of clinically relevant insights, and (3) flexibility in methods so data collection and analysis could be tailored to the intended use of findings for intervention development.	Who are the users? What is the best way to understand their unique context and specific needs? Are there existing tools and methods available that would be fit for purpose to capture user needs and preferences? How is the information going to be used? How might your approach to capturing needs and preferences be optimized for this intended use? Does data already exist (systematic reviews and qualitative research) that can help inform this step?
Step 2 (months 6-9-12) Creating intervention design objectives that capture what is unique about your intervention and reflect the specifically identified user needs and challenges the intervention needs to address.	Translate themes from the Interpretive Descriptive study into design statements and conditions for success, drawing on the data from each theme.	Helped to reframe the themes into actionable statements. Provided an evidence-based framework to underpin intervention design objectives. Ensured that user needs and preferences will continue to be reflected in the design process.	How are user needs and preferences currently expressed? Can they be used to underpin design objectives in their current form, or do they need some further refinement/transformation?
Step 2 (months 6-9-12) Creating intervention design objectives that capture what is unique about your intervention and reflect the specifically identified user needs and challenges the intervention needs to address.	Two overarching objectives for SPIN were developed from the design statements and conditions of success. It was repeatedly revised, referring to the original research question and design statements, and with input from stakeholders.	Developing 2 overarching objectives, rather than 1, helped to make explicit 2 interrelated but distinct objectives. The intermediary step of developing design objectives from the qualitative study themes ensured that the objectives represent the essence of the needs expressed by the users. Refining with input from stakeholders helped to ensure the objectives remained resonant with the SCI[c] community. Articulating these objectives at the outset was a useful reference point to keep coming back to for all later design and development phases.	What is(are) the overarching intervention objective(s)? How will you ensure your overarching objective(s) remain(s) grounded by user needs and preferences? Who might need to be involved in the development of intervention objective(s)? How will you know if your intervention objective(s) adequately capture(s) the perspectives of future users?
Step 2 (months 6-9-12) Creating intervention design objectives that capture what is unique about your intervention and reflect the specifically identified user needs and challenges the intervention needs to address.	Specific design objective were identified, drawing on the design statements and overarching intervention objectives. Working definitions were formulated with reference to original data sources and in collaborative discussions as a research team.	The development of specific design statements provided a framework to identify design requirements (system requirements) and intervention features. Investing time to develop the working definitions as a team, with reference to original data sources, was important for clarity and shared understanding.	What process will you use to generate specific design objectives from your overarching design objective(s)? Who might need to be involved in that process? What data sources do you have that you can refer to so you can refine your specific design objectives?

1. Key steps in the person-based approach	2. Methods we used to operationalize PBA[a] steps in the development of SPIN[b]	3. Strengths and opportunities of our approach	4. Questions to consider when planning this step
Step 3 (months 6 - 12) Select and refine intervention features that support the specific design objectives.	Several methods were used to support the selection and refinement of intervention features for SPIN including: (1) extracting data on intervention features from a previous systematic review on self-guided exercise interventions, (2) reviewing relevant behavioral theory, (3) undertaking a behavioral analysis, and (4) drawing on persuasive system design.	Drawing on a multiplicity of methods in this step (1) ensured an evidence-based and theoretically informed approach and (2) enabled a systematic approach to ensure intervention features were those best suited to the behavioral needs of the SPIN user. A systematic approach to identifying intervention features and mapping them back to design objectives helps to imrove the credibility of intervention design. The outcome was a clear framework for SPIN intervention design that was a useful tool to support communication with design colleagues or software developers who were then bringing SPIN to form.	What data sources are available that can help you identify potential intervention features? What midrange theories are available that can help you identify potential intervention features? Are there existing tools and methods available that would be fit for purpose to help you identify intervention features which respond to user needs and preferences? Of all the potential intervention features, which are most likely to meet the design objective(s)? Who else should be involved in this process? How might you ensure that the outcome of this process can be an accessible and usable framework for others involved in intervention development?

[a]PBA: person-based approach.

[b]SPIN: Shoulder Pain Intervention delivered over the interNet.

[c]SCI: spinal cord injury.

Ethical Considerations

Ethics approval (Auckland University of Technology Ethics Committee-AUTEC 18/263) and participant consent were received for the earlier work [33] that informed this work.

Discussion

Principal Findings

This paper has described how we applied evidence, theory, and person-based approaches in the design of a self-guided digital intervention to help pwSCI manage their shoulder pain. We have detailed the processes of applying the PBA to the design of SPIN.

This builds on Yardley and colleagues' [23] collection of work. The PBA emphasizes a detailed, qualitative understanding of users' psychosocial contexts to inform intervention design. It adds value to user-centered design by addressing factors that influence behavior change, beyond just usability. The PBA complements theory- and evidence-based frameworks, such as the "Behaviour Change Wheel" [25] by tailoring interventions to the needs and preferences of specific populations. Despite growing evidence for the use of the PBA framework in intervention design [29-31,62,63], there is little available on its operationalization. To our knowledge, the detailed reporting of each step has not been available before.

In a recent systematic review on the effectiveness of self-guided digital exercise interventions, Stavric et al [21] found that interventions with theoretical underpinning had increased congruence with the intervention features leading to significant positive results. This is supported by findings from McEwan [64] who found that theory-based interventions resulted in more consistent significant improvements in physical activity. The pwSCI and shoulder pain who will use SPIN are likely to have minimal contact with a health care professional. Therefore, successful design required an understanding of how SPIN would meet their needs and how pwSCI would use it in daily life in a self-directed way. Engaging people and evidence in intervention design is supported by a range of researchers and designers [23,34,65]. Using a person-based approach, drawing on evidence from the people who will use the intervention, to derive the behavioral strategies has been shown to be effective in a variety of settings and methods of delivery [18,29,30,66,67].

Despite acknowledgment that interventions supported by theory and evidence maximize outcomes [68,69], there remains a paucity of full intervention description or design disclosure [70-73], making it challenging to explicate the link between theory and evidence and intervention features. A key tension we encountered was the limited availability of detailed examples of how the PBA had been operationalized in practice. This required us to make interpretive decisions when translating PBA principles into design elements, often without clear guidance. Additionally, balancing adherence to the PBA's iterative, user-focused process with practical constraints such as time, resources, and access to participants posed challenges. These limitations were compounded by the fact that we were largely self-taught in the application of both the PBA and behavioral analysis frameworks.

Michie and colleagues [74] recognized the challenges and lack of clarity around the purported mechanisms by which digital interventions work during an international workshop on developing and evaluating digital interventions to promote behavior change in health. DiLiberto and colleagues [75] support the importance of "insider accounts" of intervention implementation and argue that the same transparent reporting practice should apply to intervention design. Of the 16 self-guided interventions included in our systematic review and meta-analysis conducted in the planning phase [21], only 6

provided any reference to methods used to plan, design, and develop them [19,76-81]. Of these, there was little supporting detail on how the design was carried out and none of the included studies reported exploring the behavioral needs of the users before designing the intervention. Future researchers might benefit from greater transparency and reporting of the design phase, including more practical examples of operationalizing person-based and behavioral approaches. Considerations for mitigating these challenges include allocating sufficient time and resources for user involvement beyond the planning stage, documenting key design decisions, and seeking opportunities for peer collaboration to support methodological alignment and confidence.

Strengths

We have shown commitment to providing a robust and transparent process in the operationalization of the design phase of SPIN drawing on the PBA approach. This process included explicitly addressing the identified behavioral needs of the users and kept these central throughout the entire design process. The design of SPIN has demonstrated how we used evidence (from existing literature and from a previous Interpretive Descriptive qualitative study) and theory (from behavioral analysis, "Behaviour Change Wheel," and COM-B) to enhance the person-based process. This explicit and thorough process of planning and designing SPIN has provided a blueprint for intervention development when using PBA. It also addresses many of the limitations in the reporting on the development processes for existing self-guided digital exercise and physical activity-related interventions.

Limitations

Our operationalization of the PBA design phase reflects our interpretation of the PBA steps through available readings. We acknowledge there may be other perspectives and understandings. However, we believe that it is important to make our experiences visible to build on previous work and support future intervention design. Similarly, we relied on literature and online course instruction for support when we conducted the behavioral analysis using the COM-B. Being self-taught in both the PBA and behavioral analyses may mean that some aspects of our approach are not consistent with the original intent of these approaches. However, this is perhaps an artifact of the knowledge mobilization process, where the application of knowledge can change as knowledge changes hands. By offering transparency in our process, we hope that people can draw their own conclusions regarding the robustness of our approach. The design of SPIN did not include a logic model. Logic models typically include the main intervention components, how they relate to one another, which are meant to produce which effect, and include processes and expected outcomes. However, we did not believe a logic model would have been pragmatically useful as they assume causal relationships which may have restricted our thinking about solutions [82]. Our development process was underpinned by relationship building and community interaction, both of which are complex and require flexibility [83,84].

Future Steps of SPIN Using the PBA

With the proposed intervention features selected, SPIN wireframes have been constructed. Wireframes are images or screenshots that show how screens of a website or app are structured and how content is arranged. These have provided a visual representation of the product and an opportunity to comment on content, features, and organization without getting distracted by aesthetics. Further participant consultation and design refinement have occurred. Frontend and backend software programming will occur at a later phase. Reporting of these stages will follow in a subsequent publication.

Conclusion

The design of SPIN has incorporated a deep understanding of the users' needs and best available evidence by drawing on the PBA design process to maximize chances of engagement and outcomes. This paper has made visible the operationalization of each of the phases and can act as a blueprint to provide guidance to future researchers when using this approach.

Authors' Contributions

Conceptualization: VS, NMK
Methodology: VS, NLS, NMK
Supervision: NLS, NMK
Visualization: VS
Writing – original draft: VS
Writing – review & editing: VS, NLS, NMK

Conflicts of Interest

None declared.

Multimedia Appendix 1
Systematic review data collection template.
[XLSX File, 25 KB - mhealth_v14i1e66678_app1.xlsx]

References

1. Dalyan M, Cardenas DD, Gerard B. Upper extremity pain after spinal cord injury. Spinal Cord 1999 Mar;37(3):191-195. [doi: 10.1038/sj.sc.3100802] [Medline: 10213328]
2. Curtis KA, Drysdale GA, Lanza RD, Kolber M, Vitolo RS, West R. Shoulder pain in wheelchair users with tetraplegia and paraplegia. Arch Phys Med Rehabil 1999 Apr;80(4):453-457. [doi: 10.1016/s0003-9993(99)90285-x] [Medline: 10206610]
3. Gutierrez DD, Thompson L, Kemp B, Mulroy SJ, Physical Therapy Clinical Research Network, Rehabilitation Research and Training Center on Aging-Related Changes in Impairment for Persons Living with Physical Disabilities. The relationship of shoulder pain intensity to quality of life, physical activity, and community participation in persons with paraplegia. J Spinal Cord Med 2007;30(3):251-255. [doi: 10.1080/10790268.2007.11753933] [Medline: 17684891]
4. McCasland LD, Budiman-Mak E, Weaver FM, Adams E, Miskevics S. Shoulder pain in the traumatically injured spinal cord patient: evaluation of risk factors and function. J Clin Rheumatol 2006 Aug;12(4):179-186. [doi: 10.1097/01.rhu.0000230532.54403.25] [Medline: 16891921]
5. Sie IH, Waters RL, Adkins RH, Gellman H. Upper extremity pain in the postrehabilitation spinal cord injured patient. Arch Phys Med Rehabil 1992 Jan;73(1):44-48. [doi: 10.5555/uri:pii:000399939290225L] [Medline: 1729973]
6. Coulter EH, McLean AN, Hasler JP, Allan DB, McFadyen A, Paul L. The effectiveness and satisfaction of web-based physiotherapy in people with spinal cord injury: a pilot randomised controlled trial. Spinal Cord 2017 Apr;55(4):383-389. [doi: 10.1038/sc.2016.125] [Medline: 27596027]
7. Best KL, Bourassa S, Sweet SN, Routhier F. Expert consensus for a digital peer-led approach to improving physical activity among individuals with spinal cord injury who use manual wheelchairs. J Spinal Cord Med 2023 Jan 2;46(1):53-61. [doi: 10.1080/10790268.2021.1986308]
8. Baehr LA, Hiremath SV, Bruneau M Jr, et al. Effect of Tele-exercise to Promote Empowered Movement for Individuals With Spinal Cord Injury (TEEMS) program on physical activity determinants and behavior: a mixed methods assessment. Arch Phys Med Rehabil 2024 Jan;105(1):101-111. [doi: 10.1016/j.apmr.2023.08.019] [Medline: 37678447]
9. Carey RL, Le H, Coffman DL, et al. mHealth-based just-in-time adaptive intervention to improve the physical activity levels of individuals with spinal cord injury: protocol for a randomized controlled trial. JMIR Res Protoc 2024 Jun 28;13:e57699. [doi: 10.2196/57699] [Medline: 38941145]
10. Davies CA, Spence JC, Vandelanotte C, Caperchione CM, Mummery WK. Meta-analysis of internet-delivered interventions to increase physical activity levels. Int J Behav Nutr Phys Act 2012 Apr 30;9(52):1-13. [doi: 10.1186/1479-5868-9-52] [Medline: 22546283]
11. Jahangiry L, Farhangi MA, Shab-Bidar S, Rezaei F, Pashaei T. Web-based physical activity interventions: a systematic review and meta-analysis of randomized controlled trials. Public Health (Fairfax) 2017 Nov;152:36-46. [doi: 10.1016/j.puhe.2017.06.005] [Medline: 28734170]
12. Coulter CL, Weber JM, Scarvell JM. Group physiotherapy provides similar outcomes for participants after joint replacement surgery as 1-to-1 physiotherapy: a sequential cohort study. Arch Phys Med Rehabil 2009 Oct;90(10):1727-1733. [doi: 10.1016/j.apmr.2009.04.019] [Medline: 19801063]
13. Singh G, MacGillivray M, Mills P, Adams J, Sawatzky B, Mortenson WB. Patients' perspectives on the usability of a mobile app for self-management following spinal cord injury. J Med Syst 2019 Dec 11;44(1):1-9. [doi: 10.1007/s10916-019-1487-y] [Medline: 31828440]
14. Van Straaten MG, Cloud BA, Morrow MM, Ludewig PM, Zhao KD. Effectiveness of home exercise on pain, function, and strength of manual wheelchair users with spinal cord injury: a high-dose shoulder program with telerehabilitation. Arch Phys Med Rehabil 2014 Oct;95(10):1810-1817. [doi: 10.1016/j.apmr.2014.05.004] [Medline: 24887534]
15. Bizzarini E, Chittaro L, Frezza M, et al. A mobile app for home-based exercise in spinal cord injured persons: proposal and pilot study. Digit Health 2022;8:20552076211070724. [doi: 10.1177/20552076211070724] [Medline: 35140978]
16. Brooks MA, Beaulieu JE, Severson HH, et al. Web-based therapeutic exercise resource center as a treatment for knee osteoarthritis: a prospective cohort pilot study. BMC Musculoskelet Disord 2014 May 17;15(1):2-11. [doi: 10.1186/1471-2474-15-158] [Medline: 24884547]
17. Nelligan RK, Hinman RS, Kasza J, Crofts SJC, Bennell KL. Effects of a self-directed web-based strengthening exercise and physical activity program supported by automated text messages for people with knee osteoarthritis: a randomized clinical trial. JAMA Intern Med 2021 Jun 1;181(6):776-785. [doi: 10.1001/jamainternmed.2021.0991] [Medline: 33843948]
18. Geraghty AWA, Essery R, Kirby S, et al. Internet-based vestibular rehabilitation for older adults with chronic dizziness: a randomized controlled trial in primary care. Ann Fam Med 2017 May;15(3):209-216. [doi: 10.1370/afm.2070] [Medline: 28483885]
19. van Vugt VA, van der Wouden JC, Essery R, et al. Internet based vestibular rehabilitation with and without physiotherapy support for adults aged 50 and older with a chronic vestibular syndrome in general practice: three armed randomised controlled trial. BMJ 2019 Nov 5;367:l5922. [doi: 10.1136/bmj.l5922] [Medline: 31690561]
20. Lozano-Lozano M, Martín-Martín L, Galiano-Castillo N, et al. Mobile health and supervised rehabilitation versus mobile health alone in breast cancer survivors: randomized controlled trial. Ann Phys Rehabil Med 2020 Jul;63(4):316-324. [doi: 10.1016/j.rehab.2019.07.007] [Medline: 31454561]

21. Stavric V, Kayes NM, Rashid U, Saywell NL. The effectiveness of self-guided digital interventions to improve physical activity and exercise outcomes for people with chronic conditions: a systematic review and meta-analysis. Front Rehabilit Sci 2022;3:1-16. [doi: 10.3389/fresc.2022.925620]
22. Stavric V, Saywell N, Kayes NM. Development of a self-guided web-based exercise intervention (SPIN) to treat shoulder pain in people living with spinal cord injury: protocol of a mixed methods study. BMJ Open 2019 Sep 17;9(9):e031012. [doi: 10.1136/bmjopen-2019-031012] [Medline: 31530616]
23. Yardley L, Morrison LG, 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. [doi: 10.2196/jmir.4055] [Medline: 25639757]
24. Michie S, Atkins L, West R. The Behaviour Change Wheel A Guide to Designing Interventions, 1st edition: Great Britain: Silverback Publishing; 2014:1003-1010 URL: https://www.behaviourchangewheel.com/ [accessed 2025-12-01]
25. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(42):42. [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
26. Michie S, Atkins L, Gainforth HL. Changing behaviour to improve clinical practice and policy. In: Novos Desafios, Novas Competências: Contributos Atuais Da Psicologia [Internet]: Braga: Axioma - Publicações da Faculdade de Filosofia; 2016:41-60. [doi: 10.17990/Axi/2016_9789726972679_041]
27. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4. [doi: 10.2196/jmir.1376] [Medline: 20164043]
28. Muller I, Kirby S, Yardley L. Understanding patient experiences of self-managing chronic dizziness: a qualitative study of booklet-based vestibular rehabilitation, with or without remote support. BMJ Open 2015 May 18;5(5):e007680. [doi: 10.1136/bmjopen-2015-007680] [Medline: 25986639]
29. Little P, Stuart B, Francis N, et al. Effects of internet-based training on antibiotic prescribing rates for acute respiratory-tract infections: a multinational, cluster, randomised, factorial, controlled trial. The Lancet 2013 Oct;382(9899):1175-1182. [doi: 10.1016/S0140-6736(13)60994-0]
30. Bruton A, Lee A, Yardley L, et al. Physiotherapy breathing retraining for asthma: a randomised controlled trial. Lancet Respir Med 2018 Jan;6(1):19-28. [doi: 10.1016/S2213-2600(17)30474-5] [Medline: 29248433]
31. Morton K, Dennison L, Bradbury K, et al. Qualitative process study to explore the perceived burdens and benefits of a digital intervention for self-managing high blood pressure in Primary Care in the UK. BMJ Open 2018 May 8;8(5):e020843. [doi: 10.1136/bmjopen-2017-020843] [Medline: 29739782]
32. Duncan E, O'Cathain A, Rousseau N, et al. Guidance for reporting intervention development studies in health research (GUIDED): an evidence-based consensus study. BMJ Open 2020 Apr 8;10(4):e033516. [doi: 10.1136/bmjopen-2019-033516] [Medline: 32273313]
33. Stavric V, Saywell NL, Kayes NM. Perceptions of a self-guided web-based exercise programme for shoulder pain after spinal cord injury: a qualitative study. Spinal Cord 2023 Apr;61(4):238-243. [doi: 10.1038/s41393-023-00877-3] [Medline: 36702921]
34. Morrison LG, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
35. Morrison L, Moss‐Morris R, Michie S, Yardley L. Optimizing engagement with I nternet‐based health behaviour change interventions: comparison of self‐assessment with and without tailored feedback using a mixed methods approach. British J Health Psychol 2014 Nov;19(4):839-855 [FREE Full text] [doi: 10.1111/bjhp.12083]
36. Morrison D, Mair FS, Chaudhuri R, et al. Details of development of the resource for adults with asthma in the RAISIN (randomized trial of an asthma internet self-management intervention) study. BMC Med Inform Decis Mak 2015 Jul 28;15(1-16):57. [doi: 10.1186/s12911-015-0177-z] [Medline: 26215651]
37. Dijkstra A. Personalization/computer-tailoring in persuasive technology: tailoring ingredients target psychological processes. Presented at: Proceedings of the Personalization in Persuasive Technology Workshop, Persuasive Technology 2016; Apr 5, 2016 URL: https://research.rug.nl/en/publications/personalizationcomputer-tailoring-in-persuasive-technology-tailor/ [accessed 2025-12-01]
38. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
39. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172. [doi: 10.2196/jmir.2258] [Medline: 23963284]
40. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

41. Horsch C, Lancee J, Beun RJ, Neerincx MA, Brinkman WP. Adherence to technology-mediated insomnia treatment: a meta-analysis, interviews, and focus groups. J Med Internet Res 2015 Sep 4;17(9):e214. [doi: 10.2196/jmir.4115] [Medline: 26341671]
42. Bossen D, Buskermolen M, Veenhof C, de Bakker D, Dekker J. Adherence to a web-based physical activity intervention for patients with knee and/or hip osteoarthritis: a mixed method study. J Med Internet Res 2013 Oct 16;15(10):e223. [doi: 10.2196/jmir.2742] [Medline: 24132044]
43. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(28):486-500. [doi: 10.17705/1CAIS.02428]
44. McClure JB, Shortreed SM, Bogart A, et al. The effect of program design on engagement with an internet-based smoking intervention: randomized factorial trial. J Med Internet Res 2013 Mar 25;15(3):e69. [doi: 10.2196/jmir.2508] [Medline: 23529377]
45. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013 Apr 15;15(4):e32. [doi: 10.2196/jmir.2283] [Medline: 23587561]
46. Hurling R, Fairley BW, Dias MB. Internet-based exercise intervention systems: are more interactive designs better? Psychol Health 2006 Dec;21(6):757-772. [doi: 10.1080/14768320600603257]
47. Xu X, Griva K, Koh M, et al. Creating a smartphone app for caregivers of children with atopic dermatitis with caregivers, health care professionals, and digital health experts: participatory co-design. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e16898. [doi: 10.2196/16898] [Medline: 33118949]
48. Willett M, Duda J, Fenton S, Gautrey C, Greig C, Rushton A. Effectiveness of behaviour change techniques in physiotherapy interventions to promote physical activity adherence in lower limb osteoarthritis patients: a systematic review. PLoS One 2019;14(7):e0219482. [doi: 10.1371/journal.pone.0219482] [Medline: 31291326]
49. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020;6:2055207620905411. [doi: 10.1177/2055207620905411] [Medline: 32128233]
50. Couper MP, Alexander GL, Zhang N, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. J Med Internet Res 2010 Nov 18;12(4):e52. [doi: 10.2196/jmir.1430] [Medline: 21087922]
51. Figueiras MJ, Neto DD. Challenges in “tailoring” adjustment: new ways of improving the response to chronic conditions. Eur Psychol 2019;24(1):1-6. [doi: 10.1027/1016-9040/a000348]
52. Lin H, Wu X. Intervention strategies for improving patient adherence to follow-up in the era of mobile information technology: a systematic review and meta-analysis. PLoS One 2014;9(8):e104266. [doi: 10.1371/journal.pone.0104266]
53. Alahäivälä T, Oinas-Kukkonen H. Understanding persuasion contexts in health gamification: A systematic analysis of gamified health behavior change support systems literature. Int J Med Inform 2016 Dec;96:62-70. [doi: 10.1016/j.ijmedinf.2016.02.006]
54. Glasgow RE, Christiansen SM, Kurz D, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. J Med Internet Res 2011 Jan 25;13(1):e9. [doi: 10.2196/jmir.1391] [Medline: 21371992]
55. Haines-Saah RJ, Kelly MT, Oliffe JL, Bottorff JL. Picture Me Smokefree: a qualitative study using social media and digital photography to engage young adults in tobacco reduction and cessation. J Med Internet Res 2015 Jan 26;17(1):e27. [doi: 10.2196/jmir.4061] [Medline: 25624064]
56. van Genugten L, Dusseldorp E, Webb TL, van Empelen P. Which combinations of techniques and modes of delivery in internet-based interventions effectively change health behavior? A meta-analysis. J Med Internet Res 2016 Jun 7;18(6):e155. [doi: 10.2196/jmir.4218] [Medline: 27268104]
57. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4. [doi: 10.2196/jmir.1792] [Medline: 22260810]
58. 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. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
59. Khadjesari Z, Murray E, Kalaitzaki E, et al. Impact and costs of incentives to reduce attrition in online trials: two randomized controlled trials. J Med Internet Res 2011 Mar 2;13(1):e26. [doi: 10.2196/jmir.1523] [Medline: 21371988]
60. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92. [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
61. Meade LB, Bearne LM, Sweeney LH, Alageel SH, Godfrey EL. Behaviour change techniques associated with adherence to prescribed exercise in patients with persistent musculoskeletal pain: systematic review. Br J Health Psychol 2019 Feb;24(1):10-30. [doi: 10.1111/bjhp.12324] [Medline: 29911311]
62. Heron N, O’Connor SR, Kee F, et al. Development of a digital lifestyle modification intervention for use after transient ischaemic attack or minor stroke: a person-based approach. Int J Environ Res Public Health 2021 May 2;18(9):4861. [doi: 10.3390/ijerph18094861] [Medline: 34063298]

63. Band R, Bradbury K, Morton K, et al. Intervention planning for a digital intervention for self-management of hypertension: a theory-, evidence- and person-based approach. Implement Sci 2017 Feb 23;12(1):25. [doi: 10.1186/s13012-017-0553-4] [Medline: 28231840]
64. McEwan D, Beauchamp MR, Kouvousis C, Ray CM, Wyrough A, Rhodes RE. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: a meta-analysis. Health Psychol Rev 2019 Mar;13(1):1-17. [doi: 10.1080/17437199.2018.1547120] [Medline: 30412685]
65. Nielsen J. Iterative user-interface design. Computer (Long Beach Calif) 1993;26(11):32-41. [doi: 10.1109/2.241424]
66. Remskar M, Atkinson MJ, Marks E, Ainsworth B. Understanding university student priorities for mental health and well-being support: a mixed-methods exploration using the person-based approach. Stress Health 2022 Oct;38(4):776-789. [doi: 10.1002/smi.3133] [Medline: 35137525]
67. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. Barriers and considerations in the design and implementation of digital behavioral interventions: qualitative analysis. J Med Internet Res 2022 Mar 30;24(3):e34301. [doi: 10.2196/34301] [Medline: 35353043]
68. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021(374):n2061. [doi: 10.1136/bmj.n2061]
69. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
70. Campbell M, Katikireddi SV, Sowden A, Thomson H. Lack of transparency in reporting narrative synthesis of quantitative data: a methodological assessment of systematic reviews. J Clin Epidemiol 2019 Jan;105:1-9. [doi: 10.1016/j.jclinepi.2018.08.019] [Medline: 30196129]
71. López-Nicolás R, López-López JA, Rubio-Aparicio M, Sánchez-Meca J. A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000-2020). Behav Res Methods 2022 Feb;54(1):334-349. [doi: 10.3758/s13428-021-01644-z] [Medline: 34173943]
72. Wells M, Williams B, Treweek S, Coyle J, Taylor J. Intervention description is not enough: evidence from an in-depth multiple case study on the untold role and impact of context in randomised controlled trials of seven complex interventions. Trials 2012 Jun 28;13(95):95. [doi: 10.1186/1745-6215-13-95] [Medline: 22742939]
73. Reynolds J, DiLiberto D, Mangham-Jefferies L, et al. The practice of "doing" evaluation: lessons learned from nine complex intervention trials in action. Implement Sci 2014 Jun 17;9(75):75. [doi: 10.1186/1748-5908-9-75] [Medline: 24935096]
74. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and Evaluating Digital Interventions to Promote Behavior Change in Health and Health Care: Recommendations Resulting From an International Workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
75. DiLiberto DD, Staedke SG, Nankya F, et al. Behind the scenes of the PRIME intervention: designing a complex intervention to improve malaria care at public health centres in Uganda. Glob Health Action 2015;8:29067. [doi: 10.3402/gha.v8.29067] [Medline: 26498744]
76. Holtdirk F, Mehnert A, Weiss M, et al. Results of the Optimune trial: a randomized controlled trial evaluating a novel Internet intervention for breast cancer survivors. PLoS One 2021;16(5):e0251276. [doi: 10.1371/journal.pone.0251276] [Medline: 33961667]
77. Kelechi TJ, Madisetti M, Prentice M, Mueller M. FOOTFIT physical activity mHealth intervention for minimally ambulatory individuals with venous leg ulcers. J Wound Ostomy Continence Nurs 2020;47(2):173-181. [doi: 10.1097/WON.0000000000000631]
78. Kwon H, Lee S, Jung EJ, et al. An mHealth management platform for patients with chronic obstructive pulmonary disease (efil breath): randomized controlled trial. JMIR Mhealth Uhealth 2018 Aug 24;6(8):e10502. [doi: 10.2196/10502] [Medline: 30143475]
79. Lee MK, Yun YH, Park HA, Lee ES, Jung KH, Noh DY. A web-based self-management exercise and diet intervention for breast cancer survivors: pilot randomized controlled trial. Int J Nurs Stud 2014 Dec;51(12):1557-1567. [doi: 10.1016/j.ijnurstu.2014.04.012] [Medline: 24856854]
80. Nasseri NN, Ghezelbash E, Zhai Y, et al. Feasibility of a smartphone app to enhance physical activity in progressive MS: a pilot randomized controlled pilot trial over three months. PeerJ 2020;8(6):e9303. [doi: 10.7717/peerj.9303] [Medline: 32612882]
81. Yuan SLK, Couto LA, Marques AP. Effects of a six-week mobile app versus paper book intervention on quality of life, symptoms, and self-care in patients with fibromyalgia: a randomized parallel trial. Braz J Phys Ther 2021;25(4):428-436. [doi: 10.1016/j.bjpt.2020.10.003] [Medline: 33248904]
82. Onyura B, Mullins H, Hamza DM. Five ways to get a grip on the shortcomings of logic models in program evaluation. Can Med Educ J 2021 Dec;12(6):96-99. [doi: 10.36834/cmej.71966] [Medline: 35003436]
83. Goldman KD, Schmalz KJ. Logic models: the picture worth ten thousand words. Health Promot Pract 2006 Jan;7(1):8-12. [doi: 10.1177/1524839905283230]
84. Lee PL. What's wrong with logic models. LCSA. URL: https://clearimpact.com/wp-content/uploads/2013/03/Whats-wrong-with-logic-models.pdf [accessed 2025-11-15]

Abbreviations

PBA: person-based approach
pwSCI: people living with spinal cord injury
SCI: spinal cord injury
SPIN: Shoulder Pain Intervention delivered over the interNet

Edited by L Buis; submitted 19.Sep.2024; peer-reviewed by C Gauthier, R Martin; revised version received 25.Sep.2025; accepted 09.Oct.2025; published 05.Jan.2026.

Please cite as:
Stavric V, Saywell NL, Kayes NM
Designing a Self-Guided Digital Intervention for Self-Management of Shoulder Pain in People Living With Spinal Cord Injury: Tutorial on Using a Person-Based Approach
JMIR Mhealth Uhealth 2026;14:e66678
URL: https://mhealth.jmir.org/2026/1/e66678
doi:10.2196/66678

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org