JMIR Publications
The Leading eHealth Publisher

Effectiveness of Two Highly Popular Apps: a Randomized Crossover Trial to Increase Physical Activity Among Inactive Adults

Paulina Bondaronek, April Slee, Fiona Hamilton

Submitted to: Journal of Medical Internet Research
on: February 18, 2024

Preprint
JMIR Publications

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on it's website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressively prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Preprint JMIR Publications

Effectiveness of Two Highly Popular Apps: a Randomized Crossover Trial to Increase Physical Activity Among Inactive Adults

Paulina Bondaronek[1] PhD; April Slee[2] PhD; Fiona Hamilton[3] PhD

[1]Institute of Health Informatics University College London London GB
[2]Institute of Clinical Trials and Methodology University College London London GB
[3]eHealth Unit, Research Department of Primary Care and Population Health University College London London GB

Corresponding Author:
Paulina Bondaronek PhD
Institute of Health Informatics
University College London
222 Euston Road
London
GB

Abstract

Background: Despite widespread use, smartphone apps for physical activity (PA) lack rigorous evaluation. This study examined the impact of two top PA apps through a crossover trial.

Objective: To assess the feasibility, acceptability, and effectiveness of two smartphone apps in increasing physical activity among inactive UK adults.

Methods: A randomized crossover trial was conducted with inactive UK smartphone users. After a 1-week baseline period, participants were allocated to two PA apps: a 7-minute workout and a Couch to 5k program. Assessments at 3 and 5 weeks included feasibility and effectiveness, using intention-to-treat analysis.

Results: 209 participants accessed the screening survey. 104 were eligible and consented, 63.5% (66/104) were enrolled and randomized. 87% completed the trial. For accelerometer-measured outcomes, there were no significant differences for mean change but 16/51 participants (31.4%) increased their time in moderate to vigorous PA (MVPA) by 20% from baseline following the introduction of the intervention (95% CI= 19.1% to 45.39). Self-reported PA outcomes showed significant increases: total time spent in PA (LSM= 32.52, p<.005), moderate PA (LSM= 113.68, p<.024), walking (LSM= 375.0, p<.007), and total PA (LSM= 489.46, p<.010). Sedentary behavior decreased (LSM= -123.23, p<.001). Exercise self-efficacy (LSM= 41.78, p<.0001) and intentions increased (LSM= 5.23, p<.0001). Lower baseline activity was associated with larger increase in PA (p< 0.03 for all measures).There were no significant differences between the two apps

Conclusions: Crossover trial is a feasible and acceptable method to study apps and can be used to accelerate the evidence generation for digital health. The two PA apps showed promising results, with an impact observed for 20% increase in MVPA, self-reported PA, intentions and exercise self-efficacy. The biggest improvements were in the participants with low baseline PA, who have the greatest unmet need. The study detected no differences between the apps. Clinical Trial: ClinicalTrials.gov NCT03565627

(JMIR Preprints 18/02/2024:57492)
DOI: https://doi.org/10.2196/preprints.57492

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Preprint JMIR Publications

Original Manuscript

Preprint
JMIR Publications

Original Paper

Effectiveness of Two Highly Popular Apps: a Randomized Crossover Trial to Increase Physical Activity Among Inactive Adults

Authors:

Paulina Bondaronek, PhD[1], April Slee, PhD[2], Fiona L. Hamilton, PhD[3]

Address

[1] Institute of Health Informatics, University College London, London
[2] Institute of Clinical Trials and Methodology, University College London, London, United Kingdom
[3] eHealth Unit, Research Department of Primary Care and Population Health, University College London, London, United Kingdom

Bondaronek Paulina: p.bondaronek@ucl.ac.uk; April Slee: april.slee.15@ucl.ac.uk; Fiona Hamilton, f.hamilton@ucl.ac.uk

Corresponding author: Paulina Bondaronek, Institute of Health Informatics, University College London, 222 Euston Rd., London NW1 2DA, UK

Preprint JMIR Publications

Abstract

Background: Despite widespread use, smartphone apps for physical activity (PA) lack rigorous evaluation. This study examined the impact of two top PA apps through a crossover trial.

Objective: To assess the feasibility, acceptability, and effectiveness of two smartphone apps in increasing physical activity among inactive UK adults.

Methods: A randomized crossover trial was conducted with inactive UK smartphone users. After a 1-week baseline period, participants were allocated to two PA apps: a 7-minute workout and a Couch to 5k program. Assessments at 3 and 5 weeks included feasibility and effectiveness, using intention-to-treat analysis. Trial registration: ClinicalTrials.gov NCT03565627.

Results: 209 participants accessed the screening survey. 104 were eligible and consented, 63.5% (66/104) were enrolled and randomized. 87% completed the trial. For accelerometer-measured outcomes, there were no significant differences for mean change but 16/51 participants (31.4%) increased their time in moderate to vigorous PA (MVPA) by 20% from baseline following the introduction of the intervention (95% CI= 19.1% to 45.39). Self-reported PA outcomes showed significant increases: total time spent in PA (LSM= 32.52, p<.005), moderate PA (LSM= 113.68, p<.024), walking (LSM= 375.0, p<.007), and total PA (LSM= 489.46, p<.010). Sedentary behavior decreased (LSM= -123.23, p<.001). Exercise self-efficacy (LSM= 41.78, p<.0001) and intentions increased (LSM= 5.23, p<.0001). Lower baseline activity was associated with larger increase in PA (p< 0.03 for all measures).There were no significant differences between the two apps.

Conclusions: Crossover trial is a feasible and acceptable method to study apps and can be used to accelerate the evidence generation for digital health. The two PA apps showed promising results, with an impact observed for 20% increase in MVPA, self-reported PA, intentions and exercise self-efficacy. The biggest improvements were in the participants with low baseline PA, who have the greatest unmet need. The study detected no differences between the apps.

Keywords: Randomized controlled trial; physical activity; mobile phone; health; mobile apps; mHealth; Behavior Change

Introduction

The increase in the availability of digital health technology interventions has been unprecedented in the last decade. However, despite the wide distribution and popularity of digital health products and services, many of them have been rapidly developed and implemented [1] with little or no formal evaluation to support their claims of impact [2].The scale of the issue is vast, with scholars pointing to “scientific regression” of the field [3].

Generating meaningful digital health evidence faces barriers such as limited time, high costs, and a lack of expertise [4]. Evaluations often rely on small, non-representative samples and seldom provide real-world effectiveness or cost-effectiveness evidence. Additionally, the predominance of observational studies with inadequate data collection and outcome metrics heightens the risk of biased results [5]. These factors increase the risk of biased conclusions.

There is a pressing need for innovative approaches to evidence generation and the establishment of new evidentiary standards within digital health and care [6]. Randomized controlled trials, considered the gold standard for assessing effectiveness, are resource-intensive, and present logistical challenges in the context of the rapidly changing nature of digital health interventions [7]. There is therefore a need to apply and test innovative approaches in digital health evaluation.

The purpose of this study was to assess the feasibility and the effects of one such innovative approach. We used a randomized crossover trial as a pragmatic solution to assess two highly popular apps for physical activity (PA) in a physically inactive sample. This methodology has previously been used to evaluate a subset of participants in a longitudinal study conducted on US iPhone users which assessed 4 different features to increase physical activity using a research app they developed [8]. The results were promising, with significant positive effects on step count for all 4 features.

In this study, we focus on PA due to its significant health implications, with inactivity recognized as a major risk factor for mortality and preventable diseases globally [9].

Modest increases in PA can lead to important health improvements, particularly when shifting from inactivity to moderate activity [10]. While numerous PA apps exist, their potential to increase PA has not been fully tapped. Billions of app downloads signal a vast opportunity for digital health to make a substantial impact on population-level PA.

We assessed two apps that were highly ranked in both app stores (iTunes and Google Play). To the authors’ best knowledge, this is the first crossover trial used in digital health conducted by independent researchers who did not develop the apps.

Aims and objectives

The aim of this study was to evaluate the feasibility and acceptability of employing a crossover trial methodology for evaluating digital health apps, and to determine their effectiveness in increasing physical activity and the determinants of PA in an inactive population. The two objectives were to:

- assess the feasibility and acceptability of a crossover trial to evaluate two popular PA apps available on the market
- assess and compare the effects of the two selected PA apps on PA and PA determinants

Methods

Study design and participants

This randomized, 2 × 2 crossover trial compared the effects of two highly popular PA apps. To ensure high quality of the study, we used the Cochrane Risk of Bias tool for randomized trials [11] as a guideline when designing the trial.

The Consolidated Standards of Reporting Trials (CONSORT) guidelines for reporting pilot and feasibility trials and eHealth trials were used to report this study's protocol [12]. Eligibility was restricted to adults identified as "moderately inactive" or "inactive" using the General Practice Physical Activity Questionnaire [13] who owned a "smartphone": iPhone (operating iOS 6.0 or newer) or Android (version 2.3.3 and up). Participants were excluded if they previously used the apps we intended to evaluate in the study; had medical conditions that required special attention when conducting PA; or if they were unwilling to use the accelerometer as per study instructions, as studies assessing digital interventions are likely to have high attrition rates. Those unable to perform basic functions relating to app usage such as downloading or navigating the app were also ineligible. Thus, only participants able and likely to adhere to study instructions were included.

The selection of the apps for the assessment in the study was based on a review and content analysis of the 65 most popular PA apps on the market [14]. Only apps that might be appropriate for those who engage in no or low PA met the criteria for the app selection for the trial. Workouts and Running programs (as described in the cited paper) were considered most appropriate for an inactive population. Apps had to be available on the two major apps stores (iTunes and Google Play); the apps were then sorted according to their behavior change potential (inclusion of behavior change techniques) and the apps with the most techniques were selected. These were App A: *7 Minute Workout Challenge* by *Fitness Guide Inc*., (Workout app); App B: *Couch to 5k* by *Public Health England (Running program).* See Appendix 1 for the detailed description of the interventions using TIDieR template. This study was approved by the University College London's Research Ethics Committee.

Summary of Patient and Public Involvement

Two representatives reviewed the study materials (participant information sheet, consent form, posters, and online study advertisements), to ensure the accessibility of the language and understandability of the content.

Enrolment, randomization and masking

Recruitment was conducted via posters in London, a recruitment website (callforparticipants.com), and social media. Potential participants were directed to a Qualtrics website for study details, eligibility screening, and consent. After consenting and baseline PA assessment via accelerometer, participants were randomized to app sequence A-B or B-A, using a computerized block randomization managed by AS, with the investigator blinded to the sequence list.

Procedures

This manuscript reports the findings of the quantitative component of this sequential mixed-methods feasibility crossover trial. The study design schema is presented in Appendix 2.

Demographic and outcome measures were collected face-to-face at the start of the study.

Participants wore a GT3X+ accelerometer (actiGraph, Pensacola, Florida) on an elasticated belt over

the right hip for 21 days (7 days each at baseline, week 3 and week 5). Participants were asked to wear the device during their waking hours except for water-based activities such as bathing.

Baseline PA data were collected over 2 consecutive weeks: baseline and post-baseline. The second week of assessments is referred to as “post-baseline” as participants may begin to change their behaviors prior to intervention as an effect of study participation. Follow up collection of measures and accelerometer data was conducted online using Qualtrics at weeks 1 (post-baseline), 3 and 5 (Appendix 3 for data collection schedule).

Each app was used by participants for two weeks. Participants received a brief instruction with a link to the first app (based on randomization sequence) and were asked to use the apps in a self-directed way with an aim to increase their PA level. No instruction for the frequency of usage was provided as this trial aimed to mimic “real-world” conditions.

At the end of the study, participants received £20 as an incentive for participation and to encourage completion of the study measures. In addition, each participant was reimbursed £3 for downloading of App A (App B had no cost to download and use).

Outcomes

Feasibility outcomes were recruitment rates and acceptability (trial completion and accelerometer compliance). Effectiveness outcomes were change in in objectively measured PA from baseline to 3 weeks follow up as quantified by: daily PA count (vertical count acceleration, CPM), MVPA, light, moderate, vigorous PA, sedentary behavior (SB), step count; proportion of participants who increased their time in MVPA by 20% from baseline; change from baseline to 3 weeks follow up in self-reported PA using the International Physical Activity Questionnaire (IPAQ) Short Form [15] and determinants of PA (Expected outcomes (EO), Exercise intentions, ESE); and the differences in change from baseline across the two apps.

Statistical analysis

As the primary goal of the study was feasibility, the target sample size of 60 participants was based on resource considerations and not formal sample size calculations. The National Institute for Health recommends a sample size of 30 in each arm as a pragmatic rule [16] for feasibility studies, and this recommendation was the main factor in selecting the sample size.

The analyses were performed using the intention-to-treat principal; all randomized participants were categorized according to their randomization order assignment and included regardless of compliance.

Accelerometer data were processed using Actilife software (version 6.13.3, actiGraph, LLC). Freedson’s cut-off points [17], i.e., the thresholds that categorize the CPM into PA intensities were used to define time spent in sedentary (0 -99 CPM), light (100 - 1951 CPM), moderate (1952 - 5724 CPM), vigorous: (5725 - 9498 CPM). These thresholds have been validated and are widely used in PA research, e.g. [18]. A minimum wear time of at least 480 min (at least 8 h per day), for at least 3 days was required to meet quality standards. This requirement has been used in previous studies, e.g. [19]. The non-wear time was excluded from analysis.

Descriptive statistics were used to report the socio-demographics and other characteristics of the participants, recruitment, retention rates, mood, and app-specific characteristics: usability, user ratings and engagement. Student’s t-test for independent samples, the Wilcoxon Rank-Sum test and the chi-squared or Fisher’s exact test were used to compare baseline characteristics and other

single-participant measures. The Wilcoxon Signed Rank rest was used to assess the difference in the baseline and post-baseline (week 1) PA measures.

We used the approach taken in recent randomized crossover trials in behavioral interventions) to guide the analysis approach [20]. The main analyses were as follows:

The difference in the intra-participant changes in behavioral and psychological outcomes from baseline to the period in which the participant was using the first assigned app (Period 1). This endpoint was assessed using a mixed model for repeated measures including fixed effects for period and baseline activity, and a random effect for app (App A or App B). This is similar to an ANOVA model that accounts for two different measurements (one from each period) for the same participant. The proportions and confidence intervals (CIs) for patients achieving 20% or greater increase in MVPA were calculated.

A secondary question of interest was whether there was any difference in PA across the apps. This difference was analyzed using a mixed model for repeated measures including baseline PA, app (App A or App B), period, and the app-by- period interaction as fixed effects.

Exploratory analyses were performed to assess the stability of the pre-intervention activity levels by obtaining measurements for both a baseline week and the following post-baseline week (week 1) before introducing the intervention.

Sensitivity analyses were performed to assess association between the baseline PA and the change in PA in the intervention period, weather (snow, rain, temperature), and Vector magnitude CPM. A post-hoc analysis of a log-transformed assessment of IPAQ and OE was performed to assess the impact of outliers.

SAS version 9.3 (Carry, NC) and R version 3.3.3 were used for analysis.

Results

Participant and trial characteristics

Screening and enrolment took part between Jan 15 and April13, 2018, with the final follow up conducted on 19 of May 2018. A total of 209 participants accessed the screening survey, 104 were eligible and consented, 63.5% (66/104) were enrolled and randomized. The recruitment ended after 66 participants were enrolled into the study and completed the baseline assessment. See Figure 1 for CONSORT participant flow diagram [12].

Figure 1: Flowchart for recruitment and retention to the crossover trial

Participant characteristics are summarized in Table 1. The mean age was 31.1 years (SD 11.4), and 42/66 (63.6%) were women. Twenty three of 66 (34.8 %) described themselves as non-White. There were no significant differences across groups for any baseline characteristics.

Table 1: Baseline characteristics

Baseline Characteristic	Statistic / Category	App A First	App B First	Total (n=66)

		(n=33)	(n=33)	
Age (years)	Mean ± SD	29.5 ± 10.4	32.7 ± 12.1	31.1 (11.4)
Gender (N,%)	Female	22 (66.7)	20 (60.6)	42 (63.6)
	Male	11 (33.3)	13 (39.4)	24 (36.4)
Ethnicity (N,%)	Asian	5 (15.2)	8 (24.2)	13 (19.7)
	Black	3 (9.1)	0 (0.0)	3 (4.6)
	Mixed	2 (6.1)	3 (9.1)	5 (7.6)
	White	22 (66.7)	21 (63.6)	43 (65.2)
	Other	1 (3.0)	1 (3.0)	2 (3.0)
Duration in the UK (yrs.)	Mean ± SD	6.67 ± 3.79	10.1 ± 13.5	3.5 (7.8)
Relationship (N,%)	Single	20 (60.6)	19 (57.6)	39 (59.1)
	In a relationship	13 (39.4)	12 (36.4)	25 (37.9)
	Separated	0 (0.0)	2 (6.1)	2 (3.0)
Education (N,%)	Postgraduate	15(42.4)	15 (45.4)	29 (43.9)
	Undergraduate	12 (36.4)	7 (21.2)	19 (28.8)
	Primary/secondary/college	7 (21.2)	11 (33.3)	18 (27.3)
Occupation (N,%)	Full-time education	9 (27.3)	10 (30.3)	19 (28.8)
	Full-time employment	16 (48.5)	12 (36.4)	28 (42.4)
	Part-time employment	3 (9.1)	4 (12.1)	7 (10.6)
	Retired	1 (3.0)	1 (3.0)	2 (3.0)
	Self-employed	0 (0.0)	2 (6.1)	2 (3.0)
	Unemployed	2 (6.1)	2 (6.1)	4 (6.1)
	Other	2 (6.1)	2 (6.1)	4 (6.1)
Household income (monthly, N,%)	Under £1,000	2 (6.1)	4 (12.1)	6 (9.1)
	£1,001 - £3,000	17 (51.5)	13 (39.4)	30 (45.5)
	>£3,000	11 (33.3)	11 (33.3)	22 (33.3)
	Not applicable	3 (9.1)	5 (15.2)	8 (12.1)
Downloaded PA apps before	(N,%)	21 (63.6)	22 (66.7)	43 (65.2)
Number of apps downloaded (N,%)	0	12 (36.4)	11 (33.3)	23 (34.9)
	1	10 (30.3)	11 (33.3)	21 (31.8)
	2	7 (21.2)	5 (15.2)	12 (18.2)
	3	2 (6.1)	4 (12.1)	6 (9.1)
	≥4	2 (6.1)	2 (6.1)	4 (6.1)
Downloaded running program-type app before	N,%	5 (15.2)	3 (9.1)	8 (12.1)
Downloaded HIIT-type app before	N,%	1 (3.0)	2 (6.1)	3 (4.6)
Used wearables before	N,%	5 (15.2)	10 (30.3)	15 (22.7)
Use wearable regularly	N,%	3 (9.1)	6 (18.2)	9 (86.4)
Main motivators for increasing PA (N,%):				
	Appearance	18 (54.6)	11(33.3)	29(43.9)

	Competence	28 (84.9)	3 (9.1)	8(12.1)
	Fitness	18 (57.6)	25 (75.8)	44 (66.7)
	Other	1(3)	0	1 (1.5)

[a] PA: Physical Activity

Feasibility and acceptability

Primary outcomes were feasibility (recruitment rates) and acceptability (trial completion and accelerometer compliance). Thirteen participants did not finish the trial: 3 withdrew and 10 participants did not adhere to the trial protocol (accelerometer wear protocol, 1 did not download the app). In total, 7.6% did not complete the trial through post-baseline assessments (95% CI 1.2% to 14.0%). See Figure 1.

Stability of effectiveness outcomes from baseline to post-baseline (prior to intervention)

There were significant increases in PA from the baseline to post-baseline week, including more vigorous PA at post-baseline (p<0.001), higher Intentions (p<0.001) and lower OE (p<0.001). No other tests of difference in PA measures or ESE were found to be significant.

Further inspection showed that the median and mean difference in vigorous PA (MET-min/week) between baseline and post-baseline were close to 0, and seventy percent of the participants had minimal changes between 480 and -480 METs. However, 10% of the participants had changes between 960 and 1920, leading to the significant difference in distributions. For context, an equivalent of 3000 MET minutes each week can be achieved by climbing the stairs for 10 minutes or running for 20 minutes on a daily basis [21].

Based on the increases in PA for some participants, the post-baseline measurement was used to calculate post-intervention changes in PA.

Impact of apps on objectively measured PA using accelerometer

The mean daily wear time for participants with valid data for baseline was 801.84 min (SD 84.12), for Period 1: 784.62 min (SD 89.63), and for Period 2: 819.76 min (SD 99.22).

The analysis of the accelerometer data is shown in Figure 2. There were no significant differences between baseline and Period 1 (change from baseline to 3 weeks follow up) using continuous measures. However, the point estimates are in the direction of increased PA.

Figure 2: The effectiveness analysis assessed using accelerometer

Relationship between baseline PA and change in PA in intervention period

The level of baseline activity affected the change observed in the intervention period. The results of the linear regressions showed that lower baseline activity was associated with larger increase in PA

(p< 0.03 for all measures, see Appendix 4).

Responder analysis

For both apps combined, in Period 1, 31.4% (16/51) increased their MVPA by >= 20% (95% CI= 19.1% to 45.39). In Period 2, 26.8% (11/41) increased their MVPA by 20% (95% CI= 14.2% to 42.9). The CIs for these results exclude 0, meaning that app intervention is highly likely to have produced increases in PA.

Impact of apps on self-reported outcomes

The analysis of the self-reported PA outcomes (IPAQ) showed a significant increase for total time spent in PA, moderate activity, walking, and total PA. Sedentary behavior decreased (Figure 3).
All psychological variables showed a significant difference. Exercise intentions and ESE increased. There was a small but significant decline in expected outcomes for PA. Results for Period 2, with the exception of vigorous PA, were similar in magnitude to Period 1.

Figure 3: The effectiveness analysis assessed using self-reported PA (International Physical Activity Questionnaire) and psychological predictors of PA.

Comparison of PA outcomes across apps

Although both groups increased PA compared to baseline, there were no differences in the objective PA outcomes between the two apps assessed (p> 0.05, for all objective measures).

Period effects

There was evidence of period effect for daily PA count (p= 0.010), wear time (p= 0.030), SB (p= 0.006), and moderate PA (p= 0.051). The largest improvement was seen with the 1st app assessed.

Sensitivity analyses results

Among the characteristics evaluated in the sensitivity analyses, the result of the analysis assessing the association between the baseline PA and the change in PA in the intervention period showed that more SB at baseline was associated with less sitting during the intervention period. In addition, a small effect of temperature in the intervention period was found with each increase in mean daily temperature increasing PA count by 3.28 units (95% CI -0.15 to 6.72, p=0.061).

Discussion

Principal Results

This randomized crossover trial demonstrated that using a crossover design is not only feasible but also well-accepted in studying digital health interventions, specifically PA apps, thereby accelerating the generation of evidence in this field. Notably, the two PA apps evaluated showed promising

outcomes, particularly in enhancing self-reported physical activity. There was a significant 20% increase in MVPA, and improvements in exercise self-efficacy and intentions. These benefits were most pronounced in participants with lower baseline levels of PA, hence have the potential of addressing a critical area of unmet need in public health. This study found no significant differences in the effectiveness between the two apps, suggesting that various PA apps might have comparable potential in promoting physical activity.

This study builds on the findings of a prior crossover trial, which demonstrated significant enhancements in physical activity levels through the use of a research app among US iPhone users [8]. The most notable gains were observed in individuals with initially low physical activity, aligning with evidence that even modest increases in activity can substantially improve health outcomes for those transitioning from inactivity to moderate activity [10].

Implications

The preliminary effects showed that 16/51 of patients increased their PA, and 13/51 decreased MVPA by 20%. It is pertinent to investigate the characteristics of those who increased MVPA so that the interventions might be tailored for those that are most responsive to the receiving them as digital health one-size-fits-all approach is unlikely to be sustainable [22].

Participants' self-reported change in PA was more optimistic than the objective measure. The discrepancy between the objective versus subjective account of participants' change has been documented e.g. [23]. We would recommend that an accelerometer device should be used in PA trials, as it is a gold standard in PA research [24].

There was a difference between baseline and post-baseline, so we strongly recommend that study protocols incorporate a no-treatment period when possible. The act of answering questions about a behavior, or knowing that the behavior is being recorded by an accelerometer may change the behavior without the introduction of the intervention, known as the mere measurement effect, reactivity of assessment [25] or Hawthorne effect [26].

Strengths

Our study has several strengths. This design is a feasible and practical method of assessing the impact of PA apps. The main advantage of this design is that it can estimate the overall effect of two apps. In addition, the difference between the two apps can also be assessed. Second, the prospective application of Risk of Bias tool [27] ensured high quality of the design. Third, participants were asked to use the apps in a self-directed manner. This means that the result of the study can approximate real-world behavior. Fourth, this study used a post-baseline design to assess the change in PA hence the results are more realistic.

Limitations

Our study has some limitations. First, while this study provides statistical analysis of the endpoints collected, there was no formal hypothesis testing, and so no type 1 error allocated in the design of the study for inferential analysis (to control the rate of false-positive results). No statistical inference should be drawn from these findings, and statistical tests should be interpreted as descriptive.

Second, the researchers relied on retrospective self-report engagement with the apps which may have been affected by the recall bias [28]. Third, there were higher levels of participants' education reported in comparison to the rest of the population in the UK [29]. However, the population in London has one of the highest levels of education in Europe, which our sample may reflect. [30]. Fourth, the time period of two weeks use of each intervention is unlikely to assess sustained behavioral effects, including habit formation.

Conclusions

This study demonstrated that a crossover trial is a feasible, acceptable, and is a pragmatic method to study the effects of PA apps. Moreover, the exploration of the potential of apps for increasing PA showed promising results, whereby psychological and behavioral outcomes changed following the introduction of the interventions.

However, the outcomes varied substantially, supporting the notion that no-size-fits-all. A future definitive trial will be modified to include consideration for the outcome measure (self-report versus accelerometer, binary versus continuous PA outcome), increasing the engagement with the apps, including EMA, and the incorporation of a no-treatment period.

Overall, this study demonstrated the value of utilizing alternative, yet high-quality and efficient methods to study health apps, as opposed to a status quo gold standard RCT. Evaluation is vital for developing the evidence base tools to realize the public health potential of digital health.

Acknowledgements

The authors are grateful to the late Professor Elizabeth Murray (EM) for co-supervising the study with FH.

Conflicts of Interest

PB provides consultancy in digital health development and evaluation
AS provides consultancy in design and analysis in evaluation
EM – none declared

Abbreviations

PA: Physical Activity
UK: United Kingdom
RCT: Randomized Controlled Trial
MVPA: Moderate to Vigorous Physical Activity
CI: Confidence Interval
LSM: Least Squares Mean
IPAQ: International Physical Activity Questionnaire
EO: Expected Outcomes
ESE: Exercise Self-Efficacy
CPM: Counts Per Minute
SB: Sedentary Behavior
MET: Metabolic Equivalent of Task
EMA: Ecological Momentary Assessment

Appendices

Multimedia Appendix 1
Description of the interventions using the TIDieR template

Multimedia Appendix 2
Study design schema

Multimedia Appendix 3
Data collection

Multimedia Appendix 4
Linear regression of the association between baseline PA and PA change during the intervention period (accelerometer data)

References

1. Gajarawala SN, Pelkowski JN. Telehealth benefits and barriers. The Journal for Nurse Practitioners. 2021;17(2):218-21.
2. Schwalbe N, Wahl B. Artificial intelligence and the future of global health. Lancet. 2020 May 16;395(10236):1579-86. PMID: 32416782. doi: 10.1016/s0140-6736(20)30226-9.
3. Istepanian RS, AlAnzi T. Mobile health (m-health): Evidence-based progress or scientific retrogression. Biomedical Information Technology: Elsevier; 2020. p. 717-33.
4. Khosla S, Tepie MF, Nagy MJ, Kafatos G, Seewald M, Marchese S, et al. The Alignment of Real-World Evidence and Digital Health: Realising the Opportunity. Therapeutic Innovation & Regulatory Science. 2021 2021/07/01;55(4):889-98. doi: 10.1007/s43441-021-00288-7.
5. Iyamu I, Gómez-Ramírez O, Xu AX, Chang H-J, Watt S, Mckee G, et al. Challenges in the development of digital public health interventions and mapped solutions: Findings from a scoping review. DIGITAL HEALTH. 2022;8:20552076221102255. PMID: 35656283. doi: 10.1177/20552076221102255.
6. Pham Q, Shaw J, Morita PP, Seto E, Stinson JN, Cafazzo JA. The Service of Research Analytics to Optimize Digital Health Evidence Generation: Multilevel Case Study. J Med Internet Res. 2019 Nov 11;21(11):e14849. PMID: 31710296. doi: 10.2196/14849.
7. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. NPJ digital medicine. 2020;3(1):1-14.
8. Shcherbina A, Hershman SG, Lazzeroni L, King AC, O'Sullivan JW, Hekler E, et al. The effect of digital physical activity interventions on daily step count: a randomized controlled crossover substudy of the MyHeart Counts Cardiovascular Health Study. The Lancet Digital Health. 2019;1(7):e344-e52.
9. Murray CJ, Aravkin AY, Zheng P, Abbafati C, Abbas KM, Abbasi-Kangevari M, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. The Lancet. 2020;396(10258):1223-49.
10. Warburton DE, Bredin SS. Health benefits of physical activity: a systematic review of current systematic reviews. Current opinion in cardiology. 2017;32(5):541-56.
11. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane

Collaboration's tool for assessing risk of bias in randomized trials. Bmj. 2011;343:d5928.
12. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomized pilot and feasibility trials. Pilot and feasibility studies. 2016;2(1):64.
13. Ahmad S, Harris T, Limb E, Kerry S, Victor C, Ekelund U, et al. Evaluation of reliability and validity of the General Practice Physical Activity Questionnaire (GPPAQ) in 60–74 year old primary care patients. BMC family practice. 2015;16(1):1-9.
14. Bondaronek P, Alkhaldi G, Slee A, Hamilton FL, Murray E. Quality of Publicly Available Physical Activity Apps: Review and Content Analysis. JMIR mHealth and uHealth. 2018;6(3).
15. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. The lancet. 2012;380(9838):247-57. PMID: 22818937.
16. Hooper R. Justifying sample size for a feasibility studyJan 16, 2023. Available from: http://www.rds-london.nihr.ac.uk/RDSLondon/media/RDSContent/files/PDFs/Justifying-Sample-Size-for-a-Feasibility-Study.pdf.
17. Freedson PS, Melanson E, Sirard J. Calibration of the Computer Science and Applications, Inc. accelerometer. Medicine and science in sports and exercise. 1998;30(5):777-81.
18. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of accelerometer wear and nonwear time classification algorithm. Medicine and science in sports and exercise. 2011;43(2):357.
19. Jefferis BJ, Sartini C, Lee I-M, Choi M, Amuzu A, Gutierrez C, et al. Adherence to physical activity guidelines in older adults, using objectively measured physical activity in a population-based study. BMC Public Health. 2014;14(1):382.
20. Brown TR, Simnad VI. A randomized crossover trial of dalfampridine extended release for effect on ambulatory activity in people with multiple sclerosis. International journal of MS care. 2016;18(4):170-6.
21. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. bmj. 2016;354:i3857.
22. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. International Journal of Behavioral Nutrition and Physical Activity. 2019;16(1):1-21.
23. Lee PH, Macfarlane DJ, Lam T, Stewart SM. Validity of the international physical activity questionnaire short form (IPAQ-SF): A systematic review. International Journal of Behavioral Nutrition and Physical Activity. 2011;8(1):115.
24. Young L, Hertzog M, Barnason S. Feasibility of Using Accelerometer Measurements to Assess Habitual Physical Activity in Rural Heart Failure Patients. Geriatrics. 2017;2(3):23.
25. Rodrigues AM, O'brien N, French DP, Glidewell L, Sniehotta FF. The question–behavior effect: Genuine effect or spurious phenomenon? A systematic review of randomized controlled trials with meta-analyses. Health Psychology. 2015;34(1):61.
26. Gail MH, Benichou J, Armitage P, Colton T. Encyclopedia of epidemiologic methods: John Wiley & Sons; 2000. ISBN: 0471866415.
27. Higgins J, editor. Revised Cochrane risk of bias tool for randomized trials (RoB 2.0). Additional considerations for cross-over trials2016.
28. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. Journal of medical Internet research. 2018;20(11):e292.
29. Office for National Statistics. 2011 Census: Key Statistics and Quick Statistics for Local Authorities in the United Kingdom. 2013 [16 Mar, 2022]; Available from: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/

bulletins/keystatisticsandquickstatisticsforlocalauthoritiesintheunitedkingdom/2013-12-04#background-notes.
30. Eurostat. Tertiary educational attainment, age group 25-64 by sex and NUTS 2 regions. 2023 [Sep 25, 2023]; Available from: https://data.europa.eu/data/datasets/icx9d4o6lsbwnm63bizg?locale=en.

Preprint JMIR Publications

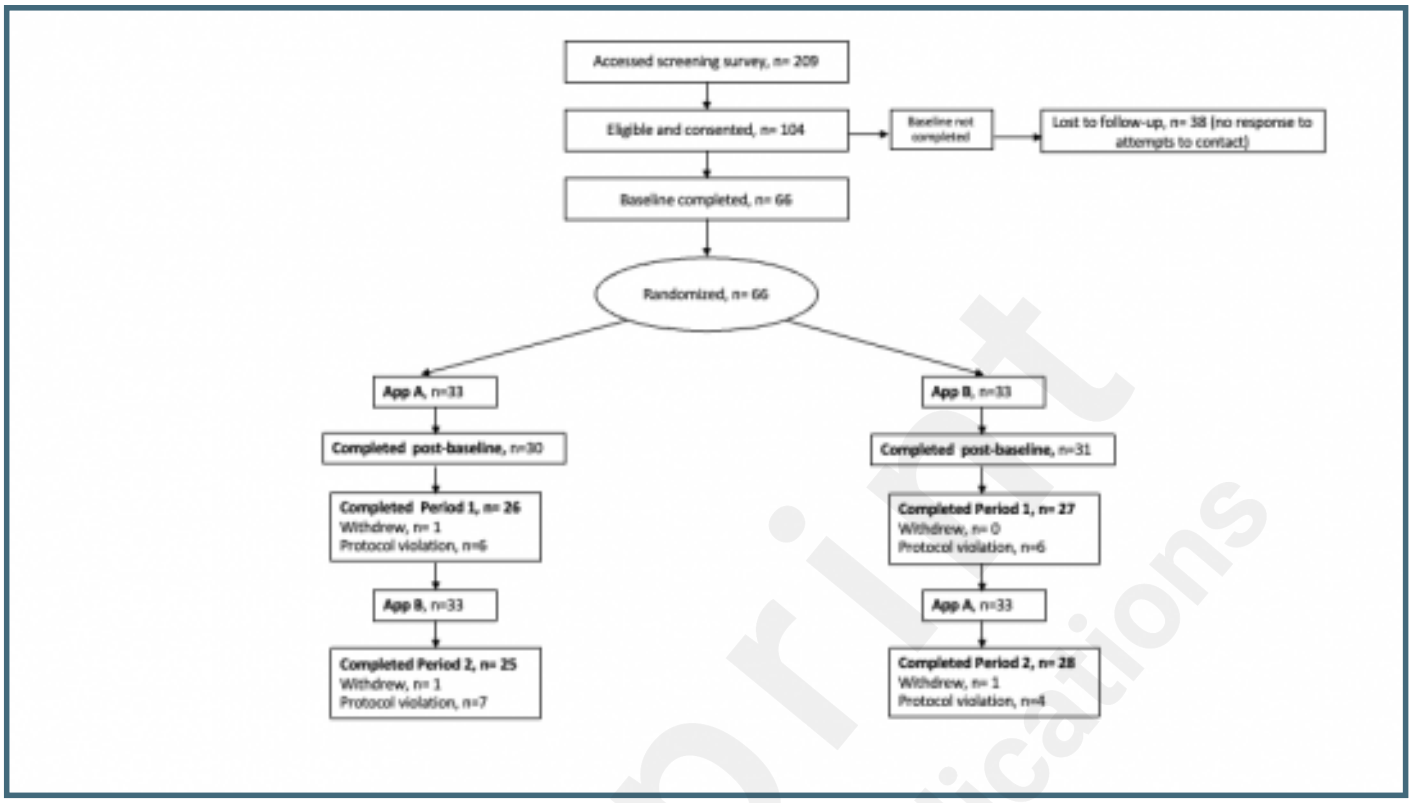
Supplementary Files

Preprint JMIR Publications

Figures

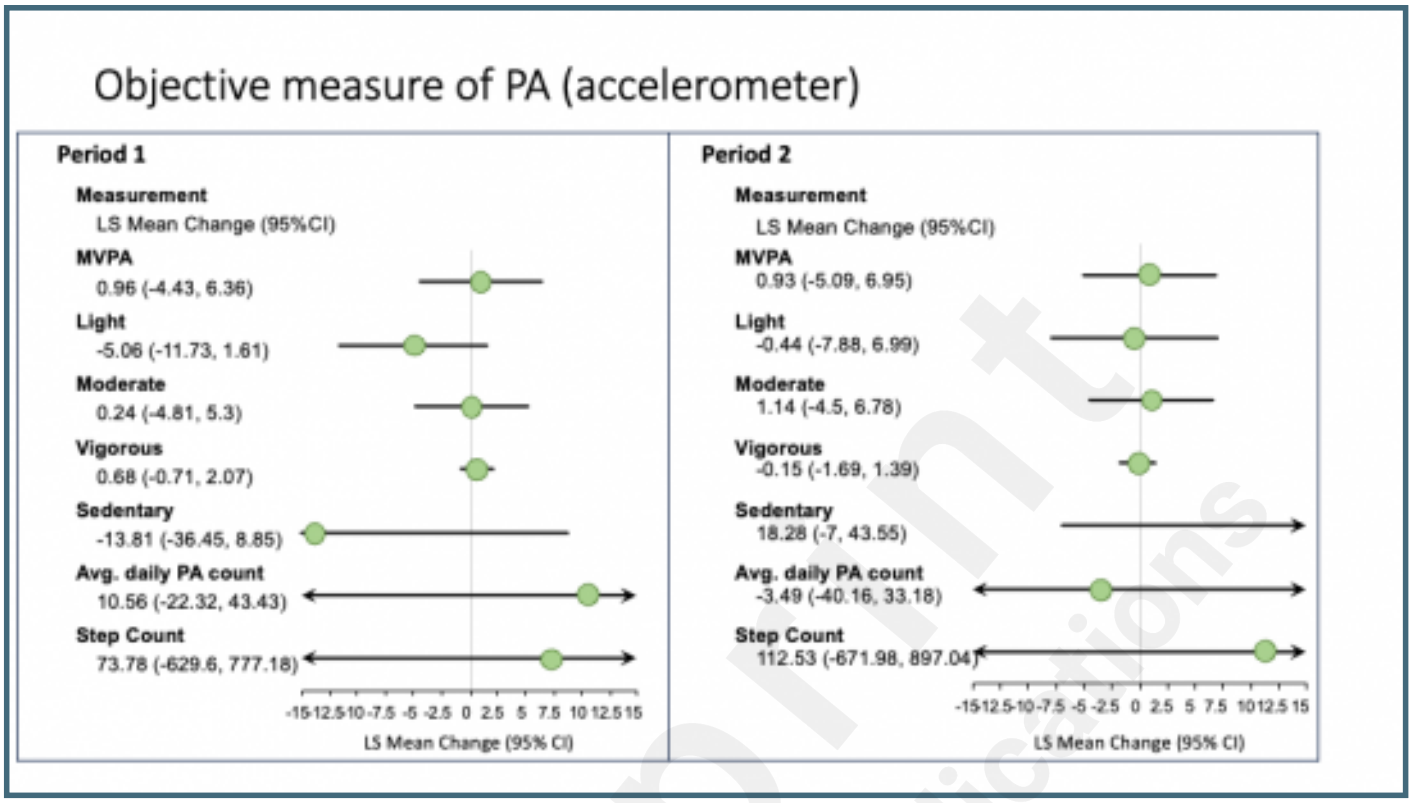
Preprint JMIR Publications

Consort flow chart.



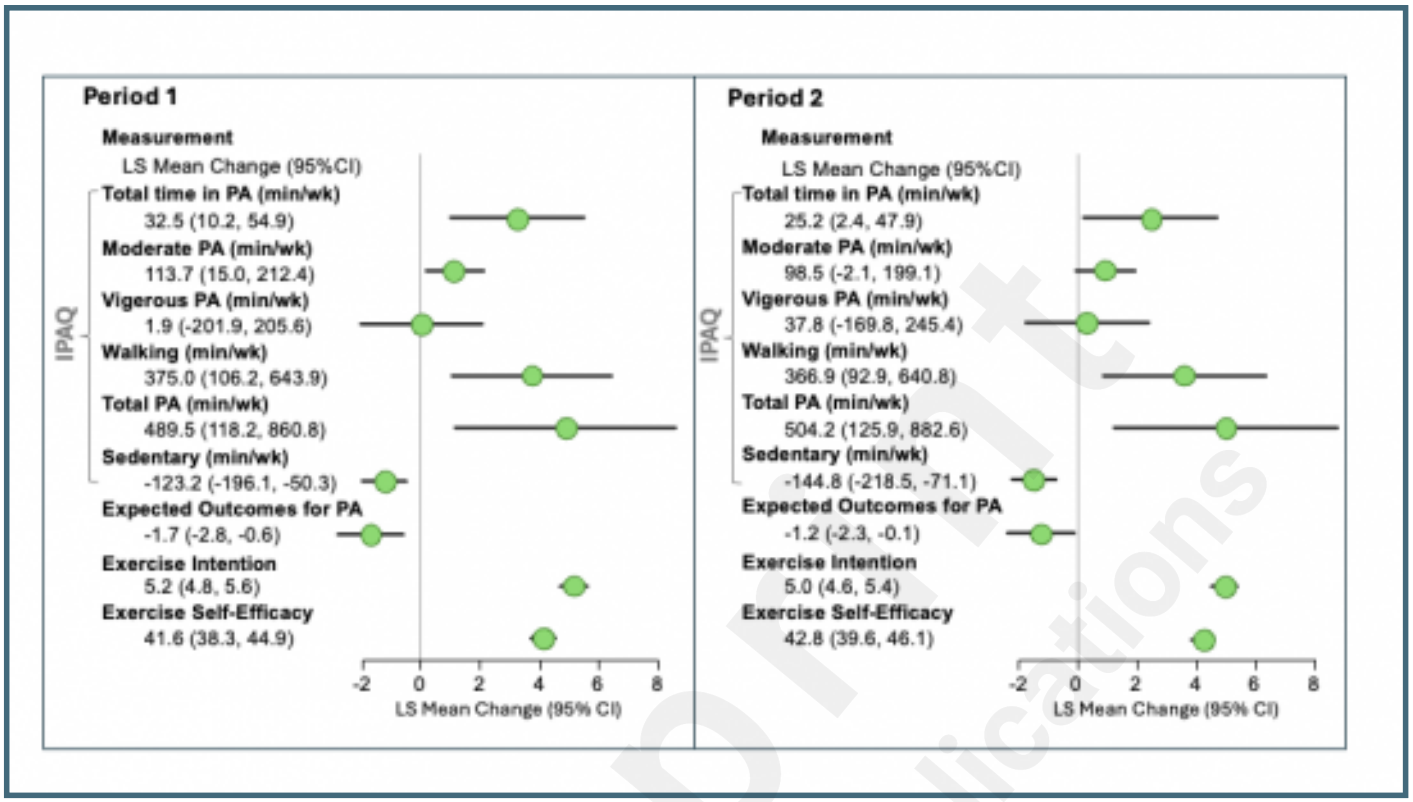
Preprint JMIR Publications

The effectiveness analysis assessed using accelerometer.



Preprint JMIR Publications

The effectiveness analysis assessed using self-reported PA (International Physical Activity Questionnaire) and psychological predictors of PA.



Preprint JMIR Publications

Multimedia Appendixes

Preprint JMIR Publications

Description of the interventions using the TIDieR template.
URL: http://asset.jmir.pub/assets/6af2ba00637f19c7f6cbb02789e48d35.docx

Study design schema.
URL: http://asset.jmir.pub/assets/abe3dd8dd1a4bdbdb8ee66c566dd549f.docx

Data collection.
URL: http://asset.jmir.pub/assets/b36a9ae49e70aec0b1fa3a56934d550b.docx

Linear regression of the association between baseline PA and PA change during the intervention period (accelerometer data).
URL: http://asset.jmir.pub/assets/3f70b76ced8a42785110ebdad632d349.docx

Preprint JMIR Publications