Assessing the Pragmatic Nature of mHealth Interventions Promoting Physical Activity: A Systematic Review and Meta-Analysis

Chad Stecher, Bjorn Pfisterer, Samantha M. Harden, Dana Epstein, Jakob M. Hirschmann, Kathrin Wunsch, Matthew P. Buman

Submitted to: Journal of Medical Internet Research
on: October 01, 2022

Table of Contents

Assessing the Pragmatic Nature of mHealth Interventions Promoting Physical Activity: A Systematic Review and Meta-Analysis

Chad Stecher[1] PhD; Bjorn Pfisterer[2] MSc; Samantha M. Harden[3] PhD; Dana Epstein[1] PhD; Jakob M. Hirschmann[4] MA; Kathrin Wunsch[2] PhD; Matthew P. Buman[1] PhD

[1]College of Health Solutions Arizona State University Phoenix US
[2]Institute of Sports and Sports Science Karlsruhe Institute of Technology Karlsruhe DE
[3]Department of Human Nutrition, Foods, and Exercise Virginia Tech Blacksburg US
[4]Institute of Sport Sciences Goethe University Frankfurt DE

Corresponding Author:
Chad Stecher PhD
College of Health Solutions
Arizona State University
500 N 3rd Street
Room 438
Phoenix
US

Abstract

Background: Mobile health (mHealth) applications (apps) can promote physical activity, but the pragmatic nature (i.e., how well research translates into real-world settings) of these studies is unknown.

Objective: The purpose of this review and meta-analysis is to describe the pragmatic nature of recent mHealth interventions for promoting physical activity and examine associations among study effect size and pragmatic study design choices.

Methods: PubMed, Scopus, Web of Science, and PsycINFO were searched up to April 2020. Studies were eligible if they incorporated apps as the primary intervention, were conducted in health promotion or preventive care settings, included a device–based physical activity outcome, and used randomized study designs. Studies were assessed with RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) and PRECIS-2 (Pragmatic-Explanatory Continuum Indicator Summary-2) frameworks. Study effect sizes were summarized using random effect models, and meta-regression was used to examine treatment effect heterogeneity by study characteristics (study duration, RE-AIM score, and PRECIS-2 score).

Results: Twenty-two interventions were selected. Data reporting across the RE-AIM framework was low overall (18.2%) and varied within specific dimensions (R=44.3%; E=52.7%; A=3.4%; I=10%; M=12.4%). PRECIS-2 results indicated that the majority of study designs were “equally explanatory and pragmatic” (63.6%). An overall positive treatment effect was observed (Cohen‘s d = 0.29 [95% CI 0.13 - 0.46]). Treatment effects varied by PRECIS-2 score (P<.01), with more explanatory studies producing larger treatment effects. Treatment effect sizes were homogenous across study duration and RE-AIM scores.

Conclusions: App-based mHealth physical activity studies have limited pragmatic utility and generalizability, and more pragmatic interventions observe smaller treatment effects. Future app-based studies should more comprehensively report real-world applicability and more pragmatic approaches are needed for maximal population health impact. Clinical Trial: International prospective register of systematic reviews (PROSPERO) CRD42020169102.

(JMIR Preprints 01/10/2022:43162)
DOI: https://doi.org/10.2196/preprints.43162

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

Original Manuscript

Assessing the Pragmatic Nature of mHealth Interventions Promoting Physical Activity: A Systematic Review and Meta-Analysis

Abstract

Background: Mobile health (mHealth) applications (apps) can promote physical activity, but the pragmatic nature (i.e., how well research translates into real-world settings) of these studies is unknown.

Objective: The purpose of this review and meta-analysis is to describe the pragmatic nature of recent mHealth interventions for promoting physical activity and examine associations among study effect size and pragmatic study design choices.

Methods: PubMed, Scopus, Web of Science, and PsycINFO were searched up to April 2020. Studies were eligible if they incorporated apps as the primary intervention, were conducted in health promotion or preventive care settings, included a device–based physical activity outcome, and used randomized study designs. Studies were assessed with RE-AIM (Reach, Effectiveness, Adoption, Implementation, Maintenance) and PRECIS-2 (Pragmatic-Explanatory Continuum Indicator Summary-2) frameworks. Study effect sizes were summarized using random effect models, and meta-regression was used to examine treatment effect heterogeneity by study characteristics (study duration, RE-AIM score, and PRECIS-2 score).

Results: Twenty-two interventions were selected. Data reporting across the RE-AIM framework was low overall (18.2%) and varied within specific dimensions (R=44.3%; E=52.7%; A=3.4%; I=10%; M=12.4%). PRECIS-2 results indicated that the majority of study designs were “equally explanatory and pragmatic” (63.6%). An overall positive treatment effect was observed (Cohen‘s d = 0.29 [95% CI 0.13 - 0.46]). Treatment effects varied by PRECIS-2 score (P<.01), with more explanatory studies producing larger treatment effects. Treatment effect sizes were homogenous across study duration and RE-AIM scores.

Conclusions: App-based mHealth physical activity studies have limited pragmatic utility and generalizability, and more pragmatic interventions observe smaller treatment effects. Future app-based studies should more comprehensively report real-world applicability and more pragmatic approaches are needed for maximal population health impact.

Trial Registration: International prospective register of systematic reviews (PROSPERO) CRD42020169102.

Keywords: physical activity; mobile health; RE-AIM; PRECIS-2; systematic review; meta-analysis; digital health

Introduction

Regular physical activity can combat numerous chronic conditions and is associated with reduced premature mortality [1,2]. Despite these benefits, behavioral interventions and public policy have been largely unsuccessful at promoting higher physical activity among the general population. Worldwide, 28% of individuals are currently classified as insufficiently active [3], and physical inactivity has an estimated annual health care cost of over $50 billion globally [4]. Thus, increasing physical activity across the world is an important economic and public health objective that requires scalable and pragmatic strategies [5].

Mobile health (mHealth) tools are one promising approach for improving health care delivery and scaling behavioral interventions worldwide [6,7]. Mobile application (app)-based platforms can be particularly effective at increasing interventions‘ accessibility and cost-effectiveness, and they offer the ability to tailor intervention methods for individuals‘ unique needs [8–10]. Accordingly, the use of app-based mHealth tools in health care has rapidly increased since 2008 [10,11], and several review papers have recently highlighted the important potential role of app–based interventions for improving global physical activity levels [12–14]. Additionally, app-based interventions saw a large, relative increase in publications compared to SMS text messaging, telehealth, or web-based interventions [14], making app-based interventions one of the most popular new clinical tools [15] and an important intervention approach to review to inform current and future researchers.

Despite the growth of research using app-based tools to promote physical activity, there is limited evidence that app–based interventions for increasing physical activity have been widely adopted by policymakers or integrated into clinical or other practice settings [16,17]. One potential explanation for this lack of real-world application is that this research has generally centered on internal validity (i.e., reliability or accuracy of the outcomes) over external validity (i.e., generalizability or applicability of results) [18,19]. In other words, the existing research

has emphasized explanatory approaches rather than more pragmatic study designs [20]. Explanatory studies measure if an intervention has a beneficial effect under ideal and thoroughly controlled circumstances, and therefore substantially differ from real world conditions (e.g., restrictive selection of study sample and control of intervention delivery). Pragmatic study designs can determine the effect of an intervention under more realistic conditions by maximizing external validity (e.g., broad and inclusive eligibility criteria and flexibility in intervention delivery) [20–23]. Studies are not strictly dichotomous in their design though, and instead are situated along the explanatory-pragmatic continuum [21,22,24]. Essentially, the challenge is striking the balance between a highly effective program and whether it can be integrated into practice settings. mHealth interventions have the unique advantages of leveraging automation, data-informed decision making, and other technological components that might aid in the adherence to the core elements (e.g., key ingredients or mechanism of change) while scaling out [25].

Existing systematic reviews of mHealth studies have broadly called for increased pragmatism [18,26,27], yet only one research review has specifically explored the generalizability and applicability of app-based physical activity interventions [16]. However, results were limited by insufficient reporting of external validity factors within the included studies. Thus, the review authors were not able to determine the generalizability of the findings, and recommended that future mHealth researchers better report on all study characteristics [16]. Specific study design characteristics, such as the study sample's demographics (e.g., average age and gender) and the duration of the intervention, are important dimensions to evaluate when determining the generalizability of a study's findings to the full population.

Given the continued growth of app–based physical activity interventions [14] and the lack of clarity surrounding the pragmatic nature of these approaches, we conducted a systematic review and meta-analysis of mHealth apps for physical activity promotion. Our primary aim

was to analyze the degree to which these interventions reported on the study characteristics necessary to inform generalizability and applicability, and to assess the explanatory versus pragmatic nature of these studies. Our secondary aim was to explore the association between study design characteristics (e.g. explanatory versus pragmatic, intervention duration, and participant demographics) and the observed effect sizes on participants‘ physical activity.

Methods

Protocol and Registration

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Multimedia Appendix 1) [28,29]. The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) (CRD42020169102).

Search Strategy and Study Selection

We conducted a systematic search in four electronic databases on April 4, 2020: PubMed, Scopus, Web of Science, and PsycINFO. The search combined synonyms and keywords related to an app–based mHealth intervention for promoting physical activity (Table 1; Multimedia Appendix 2). We attempted to control for language bias by using a search strategy without language restriction (i.e., no selective inclusion of trials published in English) [30]. In addition to these databases, the list of papers discussed by relevant systematic reviews [8,31–40] were examined to identify any further eligible studies.

Table 1. Search strategy used in PubMed on April 4, 2020.

Search Category	Search Term
(1) mHealth	mHealth OR mobile health OR m-health OR activity tracker OR fitness tracker OR wearable OR tablet OR personal digital assistant OR pda OR short message service OR sms OR text message OR android OR iphone OR iOS OR mobile phone OR cellphone OR cell phone OR cellular phone OR cellular telephone OR mobile telephone OR smart-phone OR smartphone OR mobile application OR mobile app
(2) Physical	physical activity OR leisure activity OR active living OR exercise OR

Activity	sport OR fitness OR motor activity OR sedentary behavior OR sedentary lifestyle OR sitting OR physical inactivity
(3) Intervention	Intervention OR trial OR program
(4) Study Design	clinical trial OR controlled trial OR controlled study OR double blind OR RCT OR pragmatic trial OR practical trial OR PCT OR ecological trial OR dynamic trial OR real-world OR real world
Combined	(1) AND (2) AND (3) AND (4)

The included studies were limited to app-based physical activity interventions that were published in a peer-reviewed journal between January 2012 to April 2020 that primarily targeted physical and at most one other behavioral outcome, and that presented quantitative outcome data. We further restricted our review to studies that collected device-based physical activity measures, as opposed to self-reported measures, because device-based measures are frequently observed to be more reliable [41,42] and because the use of physical activity monitoring devices has become more commonplace in the real world [43], demonstrating the feasibility, acceptability, and pragmatism of these intervention tools. A complete list of the eligibility criteria is presented in Table 2. We obtained additional data sources (when available) such as the study protocol, the Consolidated Standards of Reporting Trials (CONSORT) checklist, or any other publicly available information the corresponding authors provided via an email invitation to assess the Reach, Effectiveness, Adoption, Implementation, Maintenance (RE-AIM) framework for internal and external validity factors [44,45] and the Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) tool for evaluating interventions‘ pragmatism [24]. Specifically, this email contained a brief description of our study, and then asked: “In order to comprehensively evaluate the reporting of RE-AIM and PRECIS-2 criteria, we are also extracting data from study protocols and companion articles (e.g., qualitative or quantitative methods measuring implementation). Would you be willing to help us by providing these additional resources?”

Table 2. Eligibility criteria.[a]

Data type	Eligibility criteria
Population	Participants of any age participating in physical activity programs in the context of health promotion or preventive care settings were included. Studies focusing on special populations (e.g., pregnant women) or studies including participants with physical or psychological morbidities preventing them from participating in physical activity were excluded.
Intervention	Stand-alone mobile apps and web-apps exclusively designed for mobile devices; multi-component interventions (e.g., supported through brief counseling-sessions, or paired with other mHealth technologies) were included as long as the app was the primary component to the intervention; interventions that targeted more than one health behavior other than physical activity (e.g., diet, sleep, SB) were excluded; apps solely used for data collection purposes or as an appointment reminder service only were not eligible
Comparator	Active or inactive comparator arms were included; single-subject design trials were excluded
Outcome	Device–based measures of physical activity
Study Design	Randomized controlled trials (RCTs) and randomized ecologically valid research designs (i.e., practical clinical trials, RCTs); randomized pilot and feasibility studies were included

[a]SB: sedentary behavior; RCT: randomized controlled trials.

All records from the databases and supplementary searches were managed using Microsoft EndNote X9 reference manager software. After removing duplicates, we exported the records to Abstrackr for semi-automatic citation screening [46]. The relevance of titles and abstracts were independently assessed by two authors (BP,JH). Each eligible full-text was independently reviewed by two researchers (SMH,MB). Discrepancies were resolved through discussion between the screening authors. Any remaining conflicts were discussed among the other authors (CS,DE,KW,BP) until consensus was reached.

Data Collection Process

General Study Characteristics

We adapted an existing extraction template [32] to collect and summarize the general study characteristics. Specifically, we collected information about the study setting and design, study population, intervention components, outcome measures, key findings, and the statistical

analyses performed (Multimedia Appendix 3). Two authors (BP, JH) separately extracted additional quantitative data for the meta-analyses; discrepancies were resolved through discussion and consultation with a third author (SMH).

RE-AIM Evaluation and PRECIS-2 Assessment

We used the RE-AIM framework to describe the degree of reporting of study characteristics across five dimensions (i.e., Reach, Effectiveness, Adoption, Implementation, and Maintenance). The evaluation was assisted by a 31-item RE-AIM coding system used in previous research [47]. We then applied the PRECIS-2 tool to identify the pragmatic versus explanatory nature of each study. The tool is comprised of nine domains (i.e., eligibility criteria, recruitment, setting, flexibility (delivery), flexibility (adherence), follow-up, primary outcome, and primary analysis), each of which is assigned a score from 1 to 5 (1 is *very explanatory* and 5 is *very pragmatic*)[24]. In accordance with previous research [47], mean scores >3.5 were deemed *primarily pragmatic*. Values between 2.5 and 3.5 were considered to be *equally pragmatic and explanatory*, and scores <2.5 were rated as *primarily explanatory*. This tool compares an intervention with "usual care," and we followed the guidance of Loundon, et al. [24] and the PRECIS-2 toolkit published online, which both state that usual care is the primary care patients usually receive for medical advice and treatment.

While both frameworks can be applied regardless of study setting, additional modifications to these frameworks are recommended for a given setting [48]. Thus, we adapted the RE-AIM and PRECIS-2 coding sheets [49] for our setting (see Multimedia Appendix 4 for these adapted coding sheets). The final scoring by study is presented in Multimedia Appendix 5.

Quality Assessment

For each study, we also assessed the study's quality using a revised Cochrane risk-of-bias tool for randomized controlled trials (RoB 2.0) [50]. Two authors (BP,JH) independently made these

assessments, and any disagreements were resolved through discussion with a third author (SMH,DE,MB). Studies were classified as having a *low risk of bias* if all of the five assessment domains were considered low risk. Otherwise, studies were classified as having s*ome concerns* when concerns were raised in at least one of the five domains, or they were classified as having *high risk of bias* when at least one of the domains was judged to be at high risk. These categories were drawn from the original Cochrane RoB 2.0 tool [50].

Statistical Analyses

We used counts and percentages to summarize the general study characteristics and RE-AIM and PRECIS-2 scores for each study.

Meta-analyses were performed using the *meta* commands in Stata 16 [51]. We used the standardized average treatment effect in each study's primary app- or device-based physical activity outcome (i.e. minutes of moderate-to-vigorous physical activity (MVPA) or step count) to compare treatment effects across studies with different outcomes. The standardized average treatment effect (or Cohen's *d*) was calculated as the difference in the mean change in primary physical activity outcome between the intervention group and the control group divided by the pooled standard deviation of the physical activity outcome in both the intervention and control groups, with *a priori* interpretations [52] of trivial (<0.2), small (0.2 – 0.5), moderate (0.5 – 0.8), and large (>0.8) effects.

Additionally, we tested for heterogeneous treatment effects using random-effects models estimated through restricted maximum-likelihood (REML). All of the following moderating variables were log-transformed to better compare effect sizes: baseline physical activity, sample size, participants' age, participants' gender, intervention duration, Risk of Bias score, RE-AIM score, and PRECIS-2 score. Bubble plots were used to graphically examine the relationships between treatment effect size and the continuous moderating variables.

We assessed the statistical significance of treatment effect heterogeneity using Cochran's Q test

and calculating the Higgins I^2 statistic [53]. The following thresholds for the interpretation of the I^2 statistic were used: 0%-40%, 30%-60%, 50%-90%, or 75%-100% was interpreted as *not likely important*, *moderate*, *substantial*, or *considerable* heterogeneity, respectively [53].

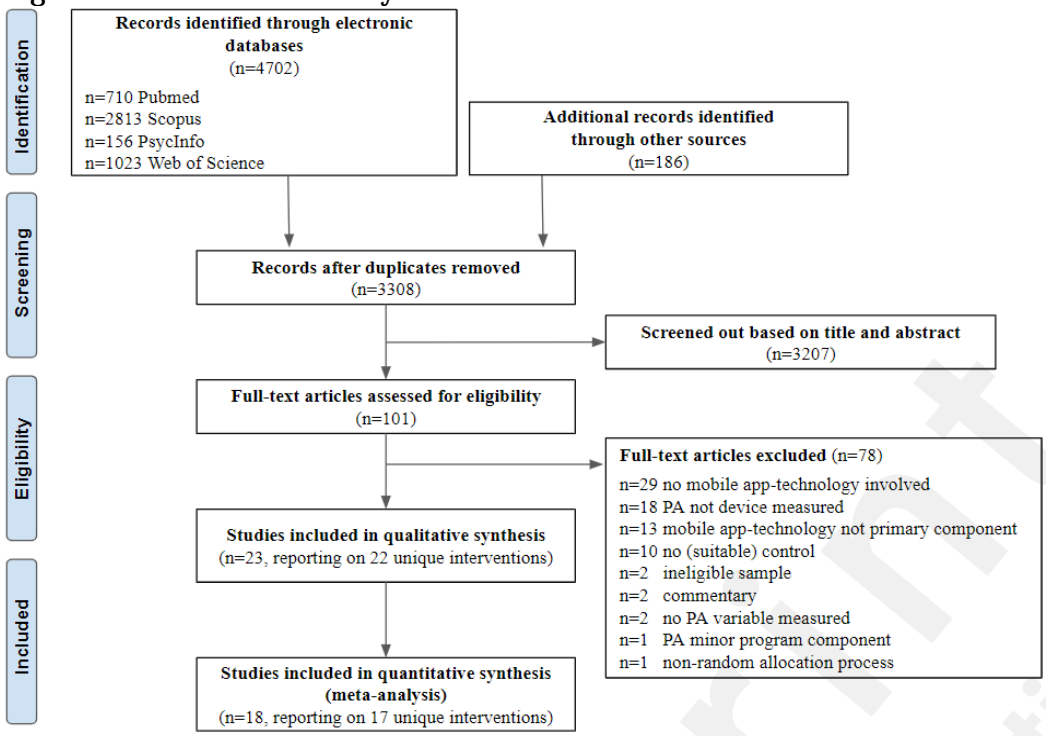
Finally, the combined impact of small-study effects and publication bias was assessed through the trim-and-fill method and performing Egger's test using the *metafor* package [54] in R version 3.6.3 [55]. The results are reported with 95% confidence intervals, and a *P* value <.05 was considered statistically significant.

Results

Study Selection

The search yielded 3,308 unique studies after duplicates were removed. We screened out 3,207 studies based on title and abstract, leaving 101 potentially relevant studies. After additional content reviews, 23 studies reporting on 22 unique interventions met the eligibility criteria for inclusion in the RE-AIM and PRECIS-2 analyses. We emailed the corresponding authors of all 23 studies to request any additional study information. We received responses from 12 of the 23 studies, and these responses either contained more information on the study (*n*=7) or simply stated that there was no additional information available (*n*=5). In total, only 17 of these 23 studies presented enough quantitative detail for inclusion in the meta-analyses. The detailed study selection process is visualized in the PRISMA flowchart (Figure 1).

Figure 1. Flowchart of study selection.[a]



[a]PA: physical activity.

Study Characteristics

All interventions were published in English between 2012 and 2020, and were conducted in ten countries, with the most common (10/22, 45.5%) based in the United States [56–65]. Of the 22 interventions, 21 used a randomized controlled trial design, of which 19 randomized subjects on an individual level and three were randomized in clusters [66–68]. One study explicitly used a pragmatic study design [69], six studies identified their trials as pilot studies [56,61,62,64,70,71], and one was classifed as a feasibility study [72]. One study performed a factorial design between multiple intervention components as part of a multiphase optimization strategy [57]. An overview of these study characteristics for each study is presented in detail in Multimedia Appendix 6.

A total of 3,555 participants were included across all 22 interventions, with sample sizes ranging from 27 to 833 (mean 161.6, SD 193.9) participants. All studies took place in a health promotion or preventive care setting, where the most common study settings were in the local

community (n=10), a university or other type of school (n=7), or a clinical care setting (n=3). Additionally, 10 interventions exclusively targeting insufficiently active individuals. Study populations varied in age and gender, with mean ages ranging from 10.6 to 61.5 years (mean of 39.6 years) and the proportion of males included across all studies was 42.8% (1,521/3,555). Two studies exclusively targeted men, and two studies included women only.

Intervention length varied from 2 weeks to 6 months (mean 60.9 days, SD 34.9). The primary app– or device-based physical activity outcome differed between interventions: the majority of studies used activity monitors or fitness trackers (77.3%, 17/22) while the rest used app-based accelerometry measures (22.7%, 5/22). All studies reported on either MVPA, daily steps, or both measures. The comparator groups received either no intervention (45.5%, 10/22), or a minimal intervention such as generic physical activity information (27.3%, 6/22), a basic app version targeting physical activity (13.6%, 3/22), a control app unrelated to physical activity (4.6%, 1/22), or a wearable activity monitor with access to its corresponding generic tracking app (9.1%, 2/22).

Six studies (27.3%, 6/22) targeted physical activity and one additional health behavior outcome (i.e., diet or sedentary behavior). In regards to the physical activity intervention strategies used among all studies, six studies provided brief in-person expert consultations (e.g., goal-setting or generic physical activity information), and one intervention included weekly telephone counseling. Most studies also used emails and text messages as physical activity reminders or to provide participants with an activity summary.

The interventions‘ apps varied greatly between studies, and consisted of both commercial products and apps designed solely for research purposes. The apps included features such as physical activity tracking and self-monitoring, feedback, goal setting, social interaction, and gamification features (see Multimedia Appendix 6 for a full list of app features by intervention).

Risk of Bias Assessment

Table 3 shows the risk of bias in the included studies. Overall, four studies showed a *low risk*, ten studies raised *some concerns,* and nine were rated *high risk.* A lack of balance across randomized study groups in terms of baseline physical activity and gender contributed to a *high risk of bias* classification for three studies, and two other studies were considered to have a *high risk of bias* for reporting on outcomes unrelated to the targeted behavior, which the authors attributed to a lack of participant engagement with the intervention's physical activity app and the intended intervention. Additionally, the majority of the studies (63.6%, 14/22) did not provide enough information to determine if data was analyzed according to their pre-specified data analysis plan, which contributed to a classification of *some concerns*.

Table 3. Risk of bias assessment based on the revised Cochrane risk-of-bias tool for randomized trials (RoB 2.0).[a]

1st author, year	Randomization bias[b]	Deviation bias[c]	Missing data bias[d]	Measurement bias[e]	Selection bias[f]	Overall
Direito, 2015	+	+	+	+	+	+
Edney, 2020	+	+	+	+	+	+
Fanning, 2017	?	?	?	+	?	−
Fukuoka, 2019	+	+	+	+	+	+
Garcia-Ortiz, 2018	+	?	?	+	+	?
Garde, 2018	+	?	?	+	?	−
Glynn, 2014	?	+	+	+	+	?
Gremaud, 2018	+	?	+	+	?	?
Harries, 2016	?	?	+	+	?	−
Hurkmans, 2018	−	?	+	+	+	−
King, 2016	+	+	+	+	?	?
Kitagawa, 2020	?	+	+	+	?	?
Leinonen, 2017	+	−	−	+	+	−
Lyons, 2017	+	+	+	+	+	+

Martin, 2015	+	+	+	+	?	?
Pope, 2020	−	?	+	+	?	−
Recio-Rodriguez, 2016	+	?	?	+	+	?
Robertson, 2018	+	−	−	+	?	−
Schade, 2020	?	?	−	+	?	−
Simons, 2018	−	?	+	+	?	−
Walsh, 2016	+	+	+	+	?	?
Zhang, 2019	+	?	+	+	?	?
Zhou, 2018	+	+	+	+	?	?

[a] + = Low risk of bias; ? = Some concerns; − = High risk of bias.
[b]Bias arising from the randomization process.
[c]Bias due to deviations from the intended intervention.
[d]Bias due to missing outcome data.
[e]Bias in selection of the reported result.

RE-AIM Evaluation

The overall rating of sufficiently reported individual RE-AIM items across all interventions was 18.2% (5.64/31, SD 2.30) (see Table 4). Reporting ranged from 2 to 11 out of the 31 RE-AIM items. The most commonly reported items were those in the Effectiveness (52.7%, 2.6/5) and Reach (44.3%, 1.8/4) dimensions. Reported data within the Maintenance categories were observed in only 12.4% (1.1/9) of the interventions, and the reporting of items in the Adoption and the Implementation dimensions were found in 3.4%, (0.3/8) and 10.0% (0.5/5) of the interventions, respectively. A summary of key findings of factors within each dimension is reported below.

Reach

Exclusion criteria commonly included health contraindications for participating in physical activity or comprised mHealth-specific requirements (e.g., specifications around technical devices). Most studies provided accurate information (i.e., either n and valid denominator, or

percentage) on the *participation rate* (72.7%, 16/22); however, only a few (13.6%, 3/22) reported the sample size in relation to the total number exposed to recruitment. The remaining trials reported only on the relation of the sample size to potentially eligible participants. Few interventions (27.2%, 6/22) adequately reported the *representativeness* of the study sample. One intervention compared their sample to eligible individuals that declined participation, and five made a comparison between their sample and their target audience. Comparisons were made on physical activity variables, and anthropometry and fitness measures.

Effectiveness

All studies reported a *measure of primary outcome* related to physical activity (per review eligibility criteria), and half of the interventions addressed a *measure of broader outcomes*. Moreover, 10 (45.5%) studies compared their physical activity-related findings to a public health goal (i.e., physical activity guidelines), yet few studies analyzed the *robustness across study subgroups* (e.g., gender, age groups). Potential explanations for physical activity-related findings were explored with *qualitative research methods* in several interventions (31.8% 7/22).

Adoption

Both non-research and research staff participation was considered, where more participation of either non-research or research staff would result in a study being less pragmatic if it exceeded the usual standard of care. However, no items were reported within the dimension Adoption-Staff. Regarding the Adoption-Setting, two studies specified *setting exclusions* (e.g., unqualified staff, unregular physical education classes). One intervention presented a valid *setting adoption rate*.

Implementation

The *delivered as intended* and the *adaptations to intervention* items were infrequently

addressed, and were mainly of technical nature (e.g., app bug or app appearance). None of the studies sufficiently reported the *cost of* intervention, meaning costs were not addressed across all levels of the intervention, or were not detailed enough (e.g., app development, technical equipment and support). The *consistency of implementation* was outlined in two trials (e.g., fidelity checks).

Maintenance

The minority of interventions (13.6%, 3/22) assessed a *≥6-month follow-up* measure; two studies reported a 6-month follow-up phase, one implemented a 9-month follow-up measure, and all of these studies stated an accurate *long-term attrition rate*. Two studies analyzed the *long-term robustness* (e.g., age, weight status). A *measure of broader outcomes* was reported in two interventions, assessing the quality of life (QoL) with the 12-Item Short-Form Health Survey.

Items within the Maintenance-Setting dimension were only addressed by three interventions, including potential *long-term adaptations* (e.g., implementing an educational app component). The sustainability of the program in a RE-AIM sense was not discussed at all.

Table 4. Inclusion of RE-AIM items across all interventions (N = 22).[a-b]

RE-AIM Dimension and Items		% (n)
Reach		**44.3%**
	Exclusion criteria	77.3 (17)
	Participation rate	72.7 (16)
	Representativeness	27.2 (6)
	Use of qualitative methods to understand reach and/or recruitment	0.0 (0)
Effectiveness		**52.7%**
	Measure of primary outcome	100.0 (22)
	Measure of broader outcomes (i.e., QoL, negative outcomes)	50.0 (11)
	Measure of robustness across subgroups	18.2 (4)
	Measure of short-term attrition	63.6 (14)
	Use of qualitative methods/data to understand outcomes	31.8 (7)
Adoption-Setting		**3.4%**
	Setting exclusions	9.1 (2)

	Setting adoption rate	4.5 (1)
	Setting representativeness	0.0 (0)
	Use of qualitative methods to understand adoption at setting level	0.0 (0)
Adoption-Staff		**0.0%**
	Staff exclusions	0.0 (0)
	Staff participation rate	0.0 (0)
	Staff representativeness	0.0 (0)
	Use of qualitative methods to understand staff participation	0.0 (0)
Implementation		**10.0%**
	Delivered as intended	22.7 (5)
	Adaptations to intervention	18.2 (4)
	Cost of intervention (time or money)	0.0 (0)
	Consistency of implementation across staff/ time/settings subgroups	9.1 (2)
	Use of qualitative methods to understand implementation	0.0 (0)
Maintenance-Individual		**9.0%**
	Measure of primary outcome at ≥6-month follow-up	13.6 (3)
	Measure of broader outcomes (i.e., QoL, negative outcomes) at follow-up	9.1 (2)
	Measure of long-term robustness across subgroups	9.1 (2)
	Measure of long-term attrition	13.6 (3)
	Use of qualitative methods to understand long-term effects	0.0 (0)
Maintenance-Setting		**3.4%**
	Program ongoing (≥6-month post-study funding)	4.5 (1)
	Long-term program adaptations	9.1 (2)
	Some discussion of sustainability of business model	0.0 (0)
	Use of qualitative methods to understand setting-level institutionalization	0.0 (0)
Overall RE-AIM		18.2%

[a]The table formatting was adapted from Burke et al.[47]
[b]RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance; QoL: quality of life.

PRECIS-2 Assessment

The overall PRECIS-2 score across all interventions is 2.93/5 (SD 0.54). Of the 22 assessed interventions, 14 interventions were categorized as *equally pragmatic and explanatory* (range 2.56-3.44), five studies were identified as being *primarily explanatory* (range 2.00-2.44), and three studies as being *primarily pragmatic* (range 3.56-4.44).

The most pragmatic dimension across all interventions was *flexibility (adherence)*, with an average score of 3.73, as demonstrated by letting participants use the app at their convenience

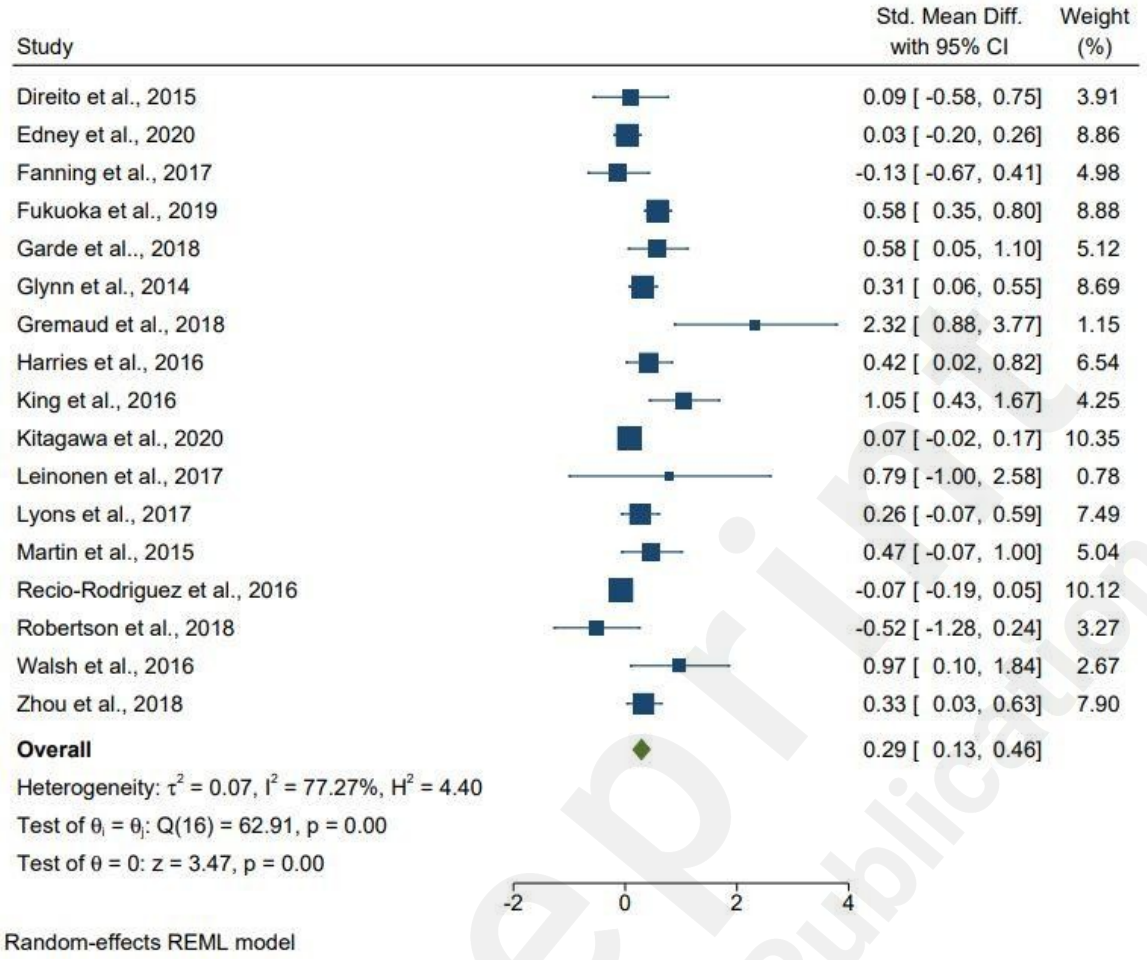
or lacking any measures to improve adherence. *Follow-up*, *organization*, and *flexibility (delivery)* appeared more explanatory with means of 2.18, 2.36, and 2.41, respectively. For example, delivery flexibility was considered more explanatory based on in-person requirements, clinician oversight, or specific app use/compliance requirements. Domains considered equally explanatory and pragmatic were *eligibility criteria*, *recruitment*, *setting*, *primary outcome*, and *primary analysis* (range 2.95-3.45). By way of example, overall, the studies in this review were equally pragmatic and explanatory in terms of eligibility criteria.

Meta-Analysis

Overall Treatment Effect

Data from only 17 interventions were extracted for this meta-analysis, because five interventions did not present complete outcome data (i.e. they did not report standard errors or 95% confidence intervals). Overall, these 17 mHealth interventions significantly improved participants' physical activity (Cohens *d* 0.29 [95% CI: 0.13 – 0.46]) (Figure 2).

Figure 2. Forest plot of standardized treatment effects on physical activity with studies weighted by the inverse of the standard error of the estimated treatment effect.[a]



[a] Std. Mean Diff.: Standardized Mean Difference; CI: Confidence Interval.

Meta-regression analyses

Meta-regression analyses revealed a statistically significant negative association between the standardized treatment effect and a study's sample size, PRECIS-2 score, and study participants' baseline physical activity (see Table 5). That is, a larger sample size, higher PRECIS-2 score (i.e., more pragmatic), and higher observed baseline physical activity levels were all associated with smaller treatment effect sizes on participants' physical activity. None of the other covariates' measures (i.e., intervention duration, participants' age, participants' gender, and RE-AIM score) were significantly related to changes in participants' physical activity.

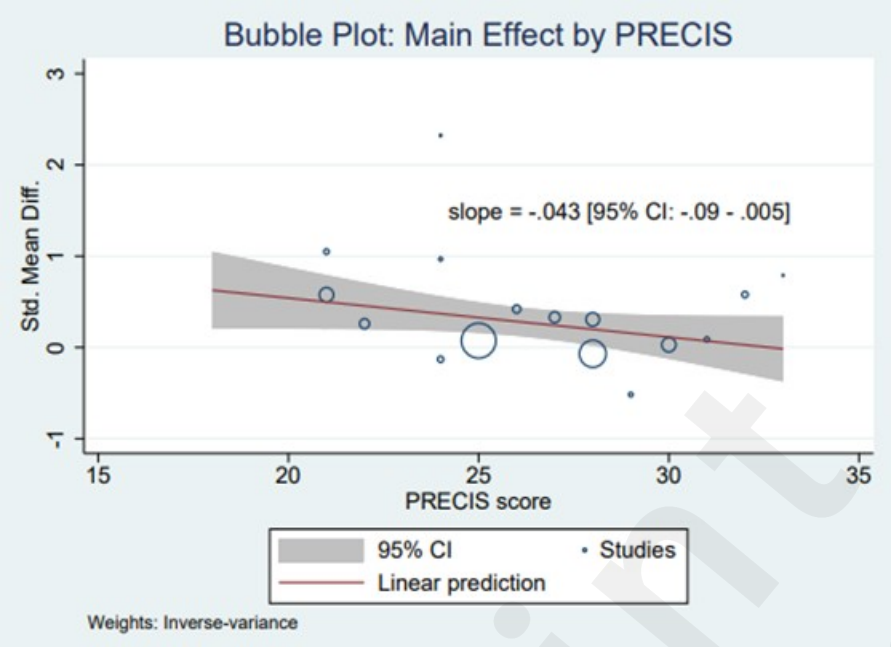
Table 5. Meta-regression results showing the interaction between study characteristics and the standardized treatment effect on physical activity.[a-b]

Covariate	Standardized mean difference
	95% confidence interval
Log(Intervention duration [days])	0.0171
	[–0.0338, 0.0680]
Log(Participant mean age [years])	–0.00296
	[–0.224, 0.218]
Log(Sample size)	–0.0616*
	[–0.111, –0.0123]
Log(Percentage male)	–0.0615
	[–0.266, 0.143]
Log(Baseline step count)	–0.420***
	[–0.637, –0.202]
Log(Baseline MVPA [minutes])	–0.199***
	[–0.288, –0.109]
Log(PRECIS-2 score)	–0.805**
	[–1.361, –0.249]
Log(RE-AIM score)	–0.0277
	[–0.177, 0.122]
Log(Risk of Bias score)	–0.199
	[–0.406, 0.0690]

[a]All covariates were log-transformed so the coefficients measure the associated change in the standardized treatment effect size from a 1% increase in the indicated variable.
[b]MVPA: moderate-to-vigorous physical activity; RE-AIM: Reach, Effectiveness, Adoption, Implementation, Maintenance; PRECIS-2: Pragmatic– Explanatory Continuum Indicator Summary-2.
*$p < .05$; **$p < .01$; ***$p < .001$.

To graphically depict the interaction between the treatment effect size and the continuous measure of a study's PRECIS-2 score, we created a bubble plot with studies represented by circles sized by the inverse of the standard error of the estimated treatment effect (Figure 3). The plot also shows the weighted linear relationship between these study characteristics and a 95% confidence interval for this estimated relationship.

Figure 3. Bubble plot of standardized treatment effect on PRECIS-2 score.[a-b]



[a]A single outlier was removed.
[b]PRECIS-2: Pragmatic– Explanatory Continuum Indicator Summary-2; Std. Mean Diff.: Standardized Mean Difference; CI: Confidence Interval.

Overall Treatment Effect Heterogeneity

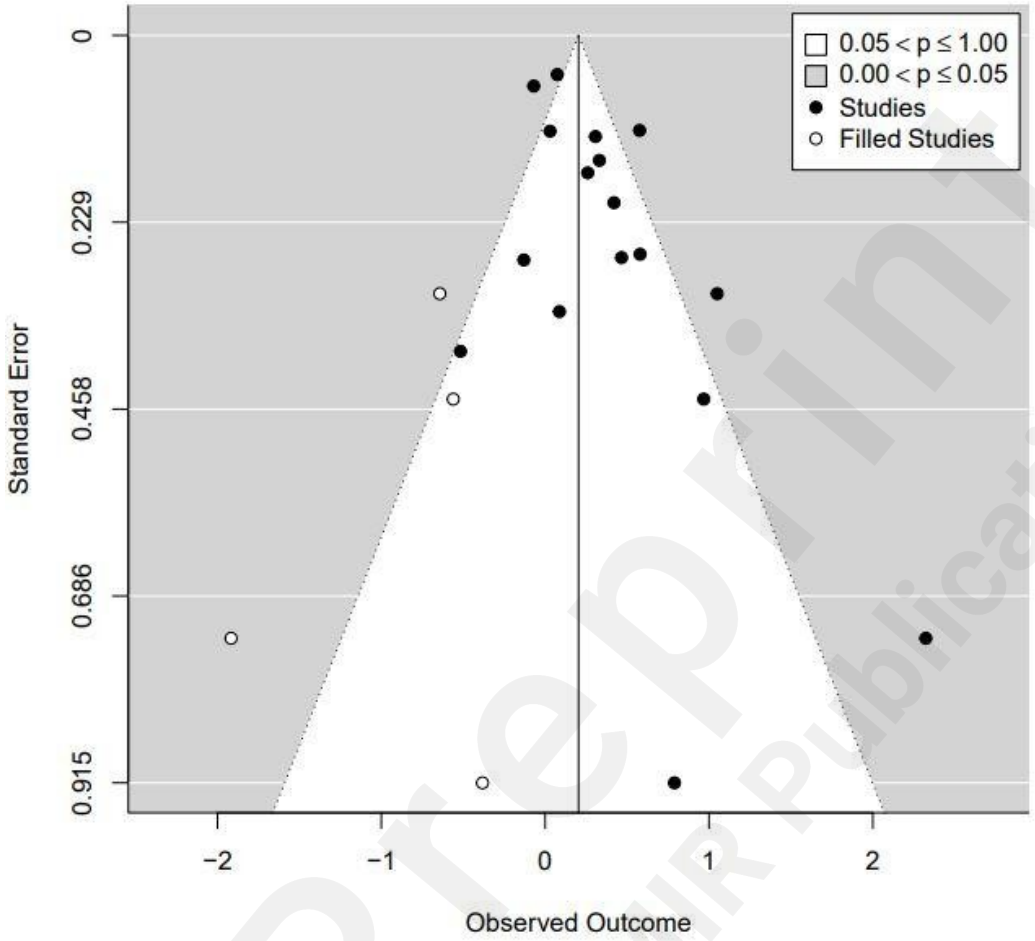
The meta-analysis showed considerable heterogeneity between studies with an I^2 value of 77.27%. The I^2 value represents the estimated percentage of variability in the results due to heterogeneity rather than due to chance.[53] The Cochran's Q test for treatment effect heterogeneity across these studies was Q(16) = 62.91, which demonstrates a statistically significant degree of heterogeneity ($p < .01$).

Analysis of Publication Bias and Small-Study Effects

We performed the trim-and-fill method to explore the potential impact of publication bias in this literature, which estimated the number of studies missing from this literature to be 4 (SE 2.80) (Figure 4). After imputing these missing studies, the overall standardized treatment effect size was slightly reduced from 0.29 (95% CI: [0.13,0.46]) to 0.20 (95% CI: [0.01,0.40]) but remained statistically significant. A high I^2 value of 83.8% indicated heterogeneity between studies remained at a considerable level after imputing these potentially missing studies. We

then carried out the Egger's test for small-study effects which reached statistical significance under most specifications (Multimedia Appendix 7).

Figure 4. Trim-and-fill funnel plot for included studies in this meta-analysis.



Discussion

Principal Findings

Among recent studies using app-based interventions to promote physical activity, we observed a significant degree of underreporting on several RE-AIM dimensions, which limits researchers' and policymakers' ability to assess the generalizability of the research results. Additionally, the interventions in this literature, in general, had more explanatory rather than pragmatic designs, which further limits our ability to forecast how successful these interventions would be on promoting physical activity if implemented among the general population. Finally, the aggregate study results show a small but significant improvement in participants' physical

activity. However, treatment effect sizes varied according to PRECIS-2 classification, where the more pragmatic trials produced smaller treatment effects on physical activity. Taken together, these findings suggest that app-based physical activity interventions would have limited efficacy for promoting physical activity if more widely scaled and adopted among the general population, suggesting that more pragmatic study designs are needed to increase the transferability from research to practice.

RE-AIM Evaluation and PRECIS-2 Assessment

RE-AIM Evaluation

Our findings build on a prior review of mHealth physical activity interventions that also observed a lack of reporting on study characteristics and research findings in this literature [16]. Without sufficient information on these important study dimensions, this prior review was unable to determine the generalizability of the research findings at that time. Our more detailed and updated review demonstrates that only small improvements in transparency and the reporting of study characteristics have been achieved in mHealth physical activity research since then.

Our finding that recent mHealth physical activity studies lack transparency builds on similar observations reported in reviews of physical activity interventions using both mHealth and other intervention tools [47,49,73]. Specifically, Blackman et al.'s 2013 review of the mHealth physical activity literature found that few studies reported on the maintenance of intervention effects and the degree of implementation fidelity [16]. Additionally, Harden et al.'s 2013 review of group-based physical activity interventions showed that external validity factors were consistently underreported [49], and Burke, et al.'s review of physical activity interventions for adults with spinal cord injuries found that several items with the Adoption and Maintenance dimensions of RE-AIM were not reported in any study, limiting the generalizability of these

studies [47].

Two specific areas of underreporting in the mHealth physical activity studies we reviewed were in the Adoption and Maintenance dimensions. The lack of reported information on the ability of healthcare providers to adopt these app-based physical activity intervention tool(s) significantly limits the willingness of clinicians and organizations to implement these new intervention approaches [16,47,49]. Additionally, none of the studies reported sufficiently on the cost of the intervention (in terms of either time or money), making it difficult to assess the benefit versus cost of these tools. Rubin and colleagues [74] have noted that prior complications experienced when integrating mHealth technologies into clinical practice has likely increased providers‘ hesitancy to adopt new mHealth strategies, so we believe that increased reporting of interventions‘ organizational requirements and cost (e.g., required staff qualifications, equipment for delivery and/or analysis, cost of acquiring the intervention tools and maintenance) would increase the applicability of this research.

PRECIS-2 Assessment

With regard to the PRECIS-2 results, our domain-specific assessments suggest that these recent studies testing app-based physical activity interventions tend to be primarily explanatory in nature. To combat a lack of app engagement, many studies employed additional text message or email reminders to re-engage participants with the interventions‘ app. These additional intervention components lowered our assessment of pragmatism as it is not clear how well these methods can be widely implemented in usual care practices. While the apps were considered relatively pragmatic in terms of their ease of accessibility, many studies also used frequent assessments, in-person intervention components, and/or brought participants into research-specific facilities, limiting their overall level of pragmatism. Importantly, the use of device-based physical activity measures did not influence the PRECIS-2 scores, as these devices are increasingly availability and integrated into usual care.

Challenges and Adaptations of RE-AIM and PRECIS-2

To address the underreporting of study characteristics, we combined the main intervention report with additional documents available online but found few additional study details through these additional sources, and thus we want to emphasize that a greater "consensus around the use of frameworks and checklists across scientific fields and journals" is still needed [47]. We also expanded the original RE-AIM framework to include a third scoring category (*inadequately/insufficiently reported),* but found that assessing this added nuance in reporting adds substantially more work to the review process. We therefore refer readers and future reviewers to the ongoing creation of domain-specific review tools [75], which will hopefully be able to strike a better balance between researcher burden and improved accuracy.

Meta-Analysis

Overall, these recent app-based physical activity interventions produced small but significant increases in participants' physical activity. This finding is in line with the results of previous reviews that also find a small and significant effect from app-based interventions for promoting physical activity [31–33,76]. Additionally, our meta-analyses found that study effect sizes were not significantly different between interventions with durations longer than 8 weeks compared to those with shorter durations (see Multimedia Appendix 7), which suggests that duration alone is not a predictor of a successful physical activity intervention and that additional approaches and intervention tools are still needed to change and maintain physical activity increases. Finally, few of these studies were able to demonstrate, or even assess, the maintenance of physical activity after the interventions were withdrawn. This finding emphasizes the need for improved understanding of physical activity habits and the maintenance of initial behavior change.

The lack of evidence for an optimal physical activity intervention duration and for the

maintenance of physical activity increases has been noted in previous reviews of the mHealth literature. Contrary to our finding, Romeo, et al. (2019) found that the most effective physical activity interventions had durations longer than 8 weeks [32]. Additionally, Schoeppe, et al.'s review of app-based health interventions showed the greatest effects among interventions up to 3 months in duration [33]. The discrepancy between our results and these reviews demonstrates the need for more evidence on the optimal intervention duration. In regards to the maintenance of intervention effects, a recent systematic review by Pradal-Cano and co-authors described the need for longer-term studies to observe the maintenance of intervention effects after the intervention components are withdrawn [76]. Among the studies reviewed by Pradal-Cano, et al., only three reported on the maintenance of intervention effects at least six months after the intervention was withdrawn and there was fixed findings on maintenance among these studies [58,66,77].

Strengths and Limitations

Adapting two complementary implementation science tools to better understand the generalizability and applicability of app–based physical activity intervention findings is a key strength of this review; however, this review is not without limitations. First, our literature search identified a relatively small number of unique interventions, which limits the power of our statistical methods. Second, the included studies significantly varied in terms of design parameters (e.g., sampling frame and intervention components) as well as methodological parameters (e.g., outcome measures). This considerable heterogeneity was identified in the meta-analyses and indicates the difficulties in synthesizing this literature. Even though we focused only on app–based physical activity interventions, most interventions incorporated additional intervention components, precluding us from isolating the individual effect of the app on physical activity. Another important limitation is that the majority of the included studies targeted adults (19/22), which limits the generalizability of our findings to physical

activity interventions among younger and older populations. Fourth, the significant degree of underreporting on study characteristics limited our ability to assess treatment moderation by individual RE-AIM dimensions, which is an important area for future research. Finally, our statistical analyses indicated the presence of a publication bias, potentially compromising the robustness of our findings. However, subsequent trim-and-fill analyses suggested that the overall treatment effect is only slightly reduced when attempting to account for these missing studies.

Conclusions

This review highlights important limitations in the mHealth literature employing app–based interventions to promote physical activity. Specifically, studies continue to underreport several key study characteristics that are necessary for determining the generalizability and scalability of these intervention approaches. Importantly, more pragmatic study designs are needed to help researchers and policymakers confidently implement app-based tools into standard care practices. Additionally, studies with different intervention durations were equally effective at increasing physical activity, which suggests that additional intervention methods and approaches are necessary for improving the maintenance and growth of initial physical activity improvements.

Acknowledgements
None.

Conflicts of Interest
None declared.

Abbreviations
CI: confidence interval
CONSORT: Consolidated Standards of Reporting Trials
mHealth: mobile health
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PCT: practical clinical trial
PRECIS-2: Pragmatic-Explanatory Continuum Indicator Summary-2
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QoL: quality of life
RCT: randomized controlled trials
RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance
SD: standard deviation

Additional files
Multimedia Appendix 1

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (doc).

Multimedia Appendix 2
Search Strategy for all Electronic Databases (pdf).

Multimedia Appendix 3

Data Extraction Form (General Study Characteristics) (pdf).

Multimedia Appendix 4

Combined Coding Sheet (Adapted RE-AIM and PRECIS-2) (pdf).

Multimedia Appendix 5

RE-AIM and PRECIS-2 scoring (pdf).

Multimedia Appendix 6

Extracted study characteristics (pdf).

Multimedia Appendix 7

Additional meta-analysis results (pdf).

References

1. Warburton DER, Bredin SSD. Health benefits of physical activity: a systematic review of current systematic reviews. Current opinion in cardiology 2017;32(5):541–556. [doi: 10.1097/HCO.0000000000000437]

2. Warburton DER, Nicol CW, Bredin SSD. Health benefits of physical activity: the evidence. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 2006;174(6):801–809. [doi: 10.1503/cmaj.051351]

3. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. The Lancet Global Health 2018;6(10):e1077–e1086. [doi: 10.1016/S2214-109X(18)30357-7]

4. Global action plan on physical activity 2018-2030: More active people for a healthier world. Geneva: World Health Organization; 2018. ISBN:978-92-4-151418-7

5. Aarons GA, Sklar M, Mustanski B, Benbow N, Brown CH. "Scaling-out" evidence-based interventions to new populations or new health care delivery systems. Implement Sci 2017 Sep 6;12(1):111. PMID:28877746

6. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, Jang H. Mobile App-Based Health Promotion Programs: A Systematic Review of the Literature. International journal of environmental research and public health 2018;15(12). [doi: 10.3390/ijerph15122838]

7. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. Journal of medical Internet research 2016;18(11):e287. [doi: 10.2196/jmir.5692]

8. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: A systematic review. Digital health 2019;5:1–50. [doi: 10.1177/2055207619839883.]

9. Kao C-K, Liebovitz DM. Consumer Mobile Health Apps: Current State, Barriers, and Future Directions. PM & R : the journal of injury, function, and rehabilitation 2017;9(5S):S106–S115. [doi: 10.1016/j.pmrj.2017.02.018]

10. Roess A. The Promise, Growth, and Reality of Mobile Health - Another Data-free Zone. The New England journal of medicine 2017;377(21):2010–2011. [doi: 10.1056/NEJMp1713180]

11. Clement J. Number of apps available in leading app stores as of 2nd quarter 2020 [Internet]. 2020. Available from: https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

12. Aitken M, Clancy B, Nass D. The growing value of digital health: Evidence and impact on human health and the healthcare system [Internet]. Parsippany, NJ: IQVIA Institute for Human Data Science; 2017. Available from: https://www.iqvia.com/insights/the-iqvia-institute/reports/the-growing-value-of-digital-health

13. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, Sohanpal R, Caton H, Griffiths CJ, Munafò MR, Taylor S, Walton RT. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ open 2016;6(10):e012447. [doi: 10.1136/bmjopen-2016-012447]

14. Müller AM, Maher CA, Vandelanotte C, Hingle M, Middelweerd A, Lopez ML, DeSmet A, Short CE, Nathan N, Hutchesson MJ, Poppe L, Woods CB, Williams SL, Wark PA. Physical Activity, Sedentary Behavior, and Diet-Related eHealth and mHealth Research: Bibliometric Analysis. Journal of medical Internet research 2018;20(4):e122. [doi: 10.2196/jmir.8954]

15. Lee J-H. Future of the smartphone for patients and healthcare providers. Healthcare informatics research The Korean Society of Medical Informatics; 2016;22(1):1–2.

16. Blackman KC, Zoellner J, Berrey LM, Alexander R, Fanning J, Hill JL, Estabrooks PA. Assessing the internal and external validity of mobile health physical activity promotion interventions: a systematic literature review using the RE-AIM framework. Journal of medical Internet research 2013;15(10):e224. [doi: 10.2196/jmir.2745]

17. Bull S. Beyond acceptability and feasibility: moving mHealth into impact. mHealth 2016;2:45. [doi: 10.21037/mhealth.2016.12.02]

18. Huebschmann AG, Leavitt IM, Glasgow RE. Making Health Research Matter: A Call to Increase Attention to External Validity. Annual review of public health 2019;40:45–63. [doi: 10.1146/annurev-publhealth-040218-043945]

19. Estabrooks PA, Gyurcsik NC. Evaluating the impact of behavioral interventions that target physical activity: issues of generalizability and public health. Psychology of Sport and Exercise 2003;4(1):41–55. [doi: 10.1016/S1469-0292(02)00016-X]

20. Glasgow RE. What does it mean to be pragmatic? Pragmatic methods, measures, and models to facilitate research translation. Health education & behavior : the official publication of the Society for Public Health Education 2013;40(3):257–265. [doi: 10.1177/1090198113486805]

21. Fritz JM, Cleland J. Effectiveness versus efficacy: more than a debate over language. The Journal of orthopaedic and sports physical therapy 2003;33(4):163–165. [doi: 10.2519/jospt.2003.33.4.163]

22. Gartlehner G, Hansen RA, Nissman D, Lohr KN, Carey TS. Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.

23. Godwin M, Ruhland L, Casson I, MacDonald S, Delva D, Birtwhistle R, Lam M, Seguin R. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. BMC medical research methodology 2003;3:28. [doi: 10.1186/1471-2288-3-28]

24. Loudon K, Treweek S, Sullivan F, Donnan P, Thorpe KE, Zwarenstein M. The PRECIS-2 tool: designing trials that are fit for purpose. BMJ (Clinical research ed) 2015;350:h2147. [doi:

10.1136/bmj.h2147]

25. Arvey SR, Fernandez ME. Identifying the core elements of effective community health worker programs: a research agenda. Am J Public Health 2012 Sep;102(9):1633–1637. PMID:22813092

26. Estabrooks PA, Brownson RC, Pronk NP. Dissemination and Implementation Science for Public Health Professionals: An Overview and Call to Action. Preventing chronic disease 2018;15:E162. [doi: 10.5888/pcd15.180525]

27. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. International journal of medical informatics 2014;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002]

28. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ (Clinical research ed) 2009;339:b2700. [doi: 10.1136/bmj.b2700]

29. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS medicine 2009;6(7):e1000097. [doi: 10.1371/journal.pmed.1000097]

30. Rothstein HR, Sutton AJ, Borenstein M. Publication Bias in Meta-Analysis. Chichester, UK: John Wiley & Sons, Ltd; 2005. [doi: 10.1002/0470870168]ISBN:978-0-470-87016-7

31. Flores Mateo G, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. Journal of medical Internet research 2015;17(11):e253. [doi: 10.2196/jmir.4836]

32. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, Crozier A, Maher C. Can Smartphone Apps Increase Physical Activity? Systematic Review and Meta-Analysis. Journal of medical Internet research 2019;21(3):e12053. [doi: 10.2196/12053]

33. Schoeppe S, Alley S, van Lippevelde W, Bray NA, Williams SL, Duncan MJ, Vandelanotte C. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. The international journal of behavioral nutrition and physical activity 2016;13(1):127. [doi: 10.1186/s12966-016-0454-y]

34. Coughlin SS, Stewart J. Use of Consumer Wearable Devices to Promote Physical Activity: A Review of Health Intervention Studies. Journal of environment and health sciences 2016;2(6). [doi: 10.15436/2378-6841.16.1123]

35. Coughlin SS, Whitehead M, Sheats JQ, Mastromonico J, Smith S. A Review of Smartphone Applications for Promoting Physical Activity. Jacobs journal of community medicine 2016;2(1).

36. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth Technologies to Influence Physical Activity and Sedentary Behaviors: Behavior Change Techniques, Systematic Review and Meta-Analysis of Randomized Controlled Trials. Annals of

behavioral medicine : a publication of the Society of Behavioral Medicine 2017;51(2):226–239. [doi: 10.1007/s12160-016-9846-0]

37. Eckerstorfer LV, Tanzer NK, Vogrincic-Haselbacher C, Kedia G, Brohmer H, Dinslaken I, Corcoran K. Key Elements of mHealth Interventions to Successfully Increase Physical Activity: Meta-Regression. JMIR mHealth and uHealth 2018;6(11):e10076. [doi: 10.2196/10076]

38. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The Effect of Physical Activity Interventions Comprising Wearables and Smartphone Applications on Physical Activity: a Systematic Review and Meta-analysis. Sports medicine - open 2018;4(1):42. [doi: 10.1186/s40798-018-0157-9]

39. Muntaner A, Vidal-Conti J, Palou P. Increasing physical activity through mobile device interventions: A systematic review. Health informatics journal 2016;22(3):451–469. [doi: 10.1177/1460458214567004]

40. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. The international journal of behavioral nutrition and physical activity 2017;14(1):105. [doi: 10.1186/s12966-017-0561-4]

41. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. The international journal of behavioral nutrition and physical activity 2008;5:56. [doi: 10.1186/1479-5868-5-56]

42. Haskell WL. Physical activity by self-report: a brief history and future issues. Journal of physical activity & health 2012;9.

43. Matthews CE, Hagströmer M, Pober DM, Bowles HR. Best practices for using physical activity monitors in population-based research. Medicine and science in sports and exercise NIH Public Access; 2012;44(1 Suppl 1):S68.

44. Lewis BA, Napolitano MA, Buman MP, Williams DM, Nigg CR. Future directions in physical activity intervention research: expanding our focus to sedentary behaviors, technology, and dissemination. Journal of behavioral medicine 2017;40(1):112–126. [doi: 10.1007/s10865-016-9797-8]

45. Gaglio B, Glasgow RE. Evaluation approaches for dissemination and implementation research. In: Brownson RC, Colditz GA, Proctor EK, editors. Dissemination and implementation research in health New York and Oxford: Oxford University Press; 2012. p. 327–356.

46. Wallace BC, Small K, Brodley CE, Lau J, Trikalinos TA. Deploying an interactive machine learning system in an evidence-based practice center. In: Luo G, Liu J, Nahapetian A, Mahoui M, editors. Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12 New York, New York, USA: ACM Press; 2012. p. 819. [doi: 10.1145/2110363.2110464]

47. Burke SM, Tomasone JR, Scime NV, Ma JK, Harden SM, Wolfe DL. Physical activity self-management interventions for adults with spinal cord injury: Part 2 – Exploring the generalizability of findings from research to practice. Psychology of Sport and Exercise 2018;37:286–295. [doi: 10.1016/j.psychsport.2018.02.013]

48. Gaglio B, Phillips SM, Heurtin-Roberts S, Sanchez MA, Glasgow RE. How pragmatic is it? Lessons learned using PRECIS and RE-AIM for determining pragmatic characteristics of research. Implementation science : IS 2014;9:96. [doi: 10.1186/s13012-014-0096-x]

49. Harden SM, Burke SM, Haile AM, Estabrooks PA. Generalizing the findings from group dynamics-based physical activity research to practice settings: what do we know? Evaluation & the health professions 2015;38(1):3–14. [doi: 10.1177/0163278713488117]

50. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, Cates CJ, Cheng H-Y, Corbett MS, Eldridge SM, Emberson JR, Hernán MA, Hopewell S, Hróbjartsson A, Junqueira DR, Jüni P, Kirkham JJ, Lasserson T, Li T, McAleenan A, Reeves BC, Shepperd S, Shrier I, Stewart LA, Tilling K, White IR, Whiting PF, Higgins JPT. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ (Clinical research ed) 2019;366:l4898. [doi: 10.1136/bmj.l4898]

51. StataCorp. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC; 2019.

52. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New York: Routledge; 1988. [doi: 10.4324/9780203771587]ISBN:978-0-203-77158-7

53. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ (Clinical research ed) 2003;327(7414):557–560. [doi: 10.1136/bmj.327.7414.557]

54. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software 2010;36(3). [doi: 10.18637/jss.v036.i03]

55. R Core Team. R: A language and environment for statistical computing [Internet]. The R Project for Statistical Computing; 2020. Available from: https://www.r-project.org/

56. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, Michos ED, Ndumele CE, Ratchford EV, Coresh J, Blaha MJ. mActive: A Randomized Clinical Trial of an Automated mHealth Intervention for Physical Activity Promotion. Journal of the American Heart Association 2015;4(11). [doi: 10.1161/JAHA.115.002239]

57. Fanning J, Roberts S, Hillman CH, Mullen SP, Ritterband L, McAuley E. A smartphone \ textquotedblapp\textquotedbl-delivered randomized factorial trial targeting physical activity in adults. Journal of behavioral medicine 2017;40(5):712–729. [doi: 10.1007/s10865-017-9838-y]

58. 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 network open 2019;2(5):e194281. [doi: 10.1001/jamanetworkopen.2019.4281]

59. Gremaud AL, Carr LJ, Simmering JE, Evans NJ, Cremer JF, Segre AM, Polgreen LA, Polgreen PM. Gamifying Accelerometer Use Increases Physical Activity Levels of Sedentary Office Workers. Journal of the American Heart Association 2018;7(13). [doi: 10.1161/JAHA.117.007735]

60. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, Banerjee B, Robinson TN, Cirimele J. 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]

61. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR mHealth and uHealth 2017;5(3):e28. [doi: 10.2196/mhealth.6967]

62. Pope ZC, Gao Z. Feasibility of smartphone application- and social media-based intervention on college students' health outcomes: A pilot randomized trial. Journal of American college health : J of ACH 2020;1–10. [doi: 10.1080/07448481.2020.1726925]

63. Schade S, Mahoney J, Spotts A, Greenauer N, Veerabhadrappa P. Pokémon Go did not increase step count or distance travelled among college students. Human Movement 2020;21(2):64–70. [doi: 10.5114/hm.2020.89916]

64. Zhang J, Jemmott III JB. Mobile App-Based Small-Group Physical Activity Intervention for Young African American Women: a Pilot Randomized Controlled Trial. Prevention science : the official journal of the Society for Prevention Research 2019;20(6):863–872. [doi: 10.1007/s11121-019-01006-4]

65. Zhou M, Fukuoka Y, Mintz Y, Goldberg K, Kaminsky P, Flowers E, Aswani A. Evaluating Machine Learning-Based Automated Personalized Daily Step Goals Delivered Through a Mobile Phone App: Randomized Controlled Trial. JMIR mHealth and uHealth 2018;6(1):e28. [doi: 10.2196/mhealth.9117]

66. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, Maher CA. A Social Networking and Gamified App to Increase Physical Activity: Cluster RCT. American journal of preventive medicine 2020;58(2):e51–e62. [doi: 10.1016/j.amepre.2019.09.009]

67. Robertson J, Macvean A, Fawkner S, Baker G, Jepson RG. Savouring our mistakes: Learning from the FitQuest project. International Journal of Child-Computer Interaction 2018;16:55–67. [doi: 10.1016/j.ijcci.2017.12.003]

68. Simons D, de Bourdeaudhuij I, Clarys P, de Cocker K, Vandelanotte C, Deforche B. Effect and Process Evaluation of a Smartphone App to Promote an Active Lifestyle in Lower Educated Working Young Adults: Cluster Randomized Controlled Trial. JMIR mHealth and uHealth 2018;6(8):e10003. [doi: 10.2196/10003]

69. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and Increasing Physical Activity Among Young People: The AIMFIT Pragmatic Randomized Controlled Trial. Journal of medical Internet research 2015;17(8):e210. [doi: 10.2196/jmir.4568]

70. Kitagawa T, Higuchi Y, Todo E, Ueda T, Ando S, Murakami T. Tailored feedback reduced prolonged sitting time and improved the health of housewives: a single-blind randomized controlled pilot study. Women & health 2020;60(2):212–223. [doi: 10.1080/03630242.2019.1616043]

71. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mHealth Intervention Using a Smartphone App to Increase Walking Behavior in Young Adults: A Pilot Study. JMIR mHealth and uHealth 2016;4(3):e109. [doi: 10.2196/mhealth.5227]

72. Leinonen A-M, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, Enwald H, Ikäheimo TM, Röning J, Keinänen-Kiukaanniemi S, Mäntysaari M, Korpelainen R, Jämsä T. Feasibility of Gamified Mobile Service Aimed at Physical Activation in Young Men: Population-Based Randomized Controlled Study (MOPO). JMIR mHealth and uHealth 2017;5(10):e146. [doi: 10.2196/mhealth.6675]

73. Galaviz KI, Harden SM, Smith E, Blackman KC, Berrey LM, Mama SK, Almeida FA, Lee RE, Estabrooks PA. Physical activity promotion in Latin American populations: a systematic review on issues of internal and external validity. The international journal of behavioral nutrition and physical activity 2014;11:77. [doi: 10.1186/1479-5868-11-77]

74. Rubin DS, Rich S, Arena R, Bond S. Leveraging technology to move more and sit less. Progress in cardiovascular diseases 2020; [doi: 10.1016/j.pcad.2020.10.007]

75. Norton WE, Loudon K, Chambers DA, Zwarenstein M. Designing provider-focused implementation trials with purpose and intent: introducing the PRECIS-2-PS tool. Implementation science : IS 2021;16(1):7. [doi: 10.1186/s13012-020-01075-y]

76. Pradal-Cano L, Lozano-Ruiz C, Pereyra-Rodríguez JJ, Saigí-Rubió F, Bach-Faig A, Esquius L, Medina FX, Aguilar-Martínez A. Using Mobile Applications to Increase Physical Activity: A Systematic Review. International journal of environmental research and public health 2020;17(21). [doi: 10.3390/ijerph17218238]

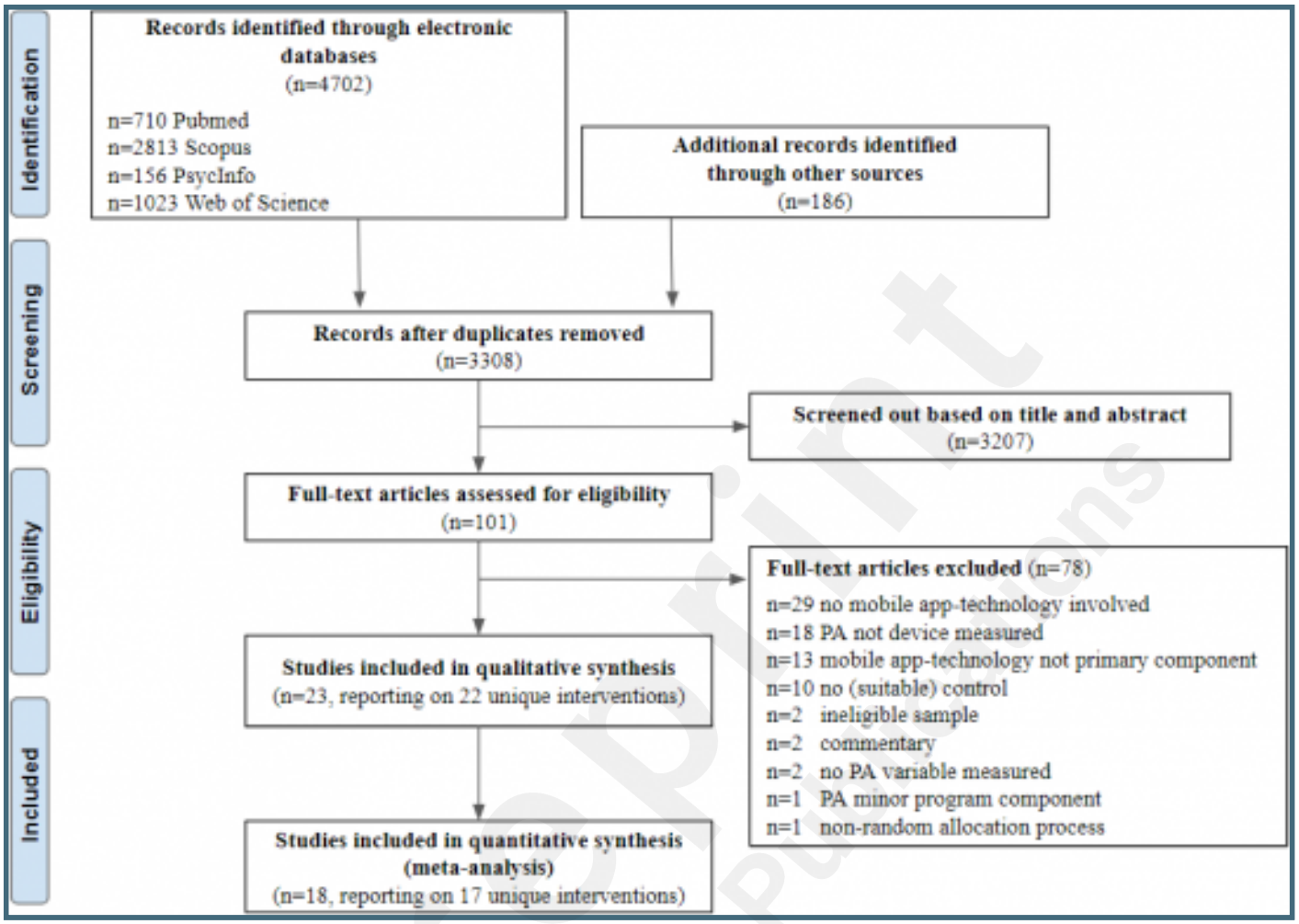
77. Garcia-Ortiz L, Recio-Rodriguez JI, Agudo-Conde C, Patino-Alonso MC, Maderuelo-Fernandez J-A, Repiso Gento I, Puigdomenech Puig E, Gonzalez-Viejo N, Arietaleanizbeaskoa MS, Schmolling-Guinovart Y, Gomez-Marcos MA, Rodriguez-Sanchez E. Long-Term Effectiveness of a Smartphone App for Improving Healthy Lifestyles in General Population in Primary Care: Randomized Controlled Trial (Evident II Study). JMIR mHealth and uHealth 2018;6(4):e107. [doi: 10.2196/mhealth.9218]

Supplementary Files

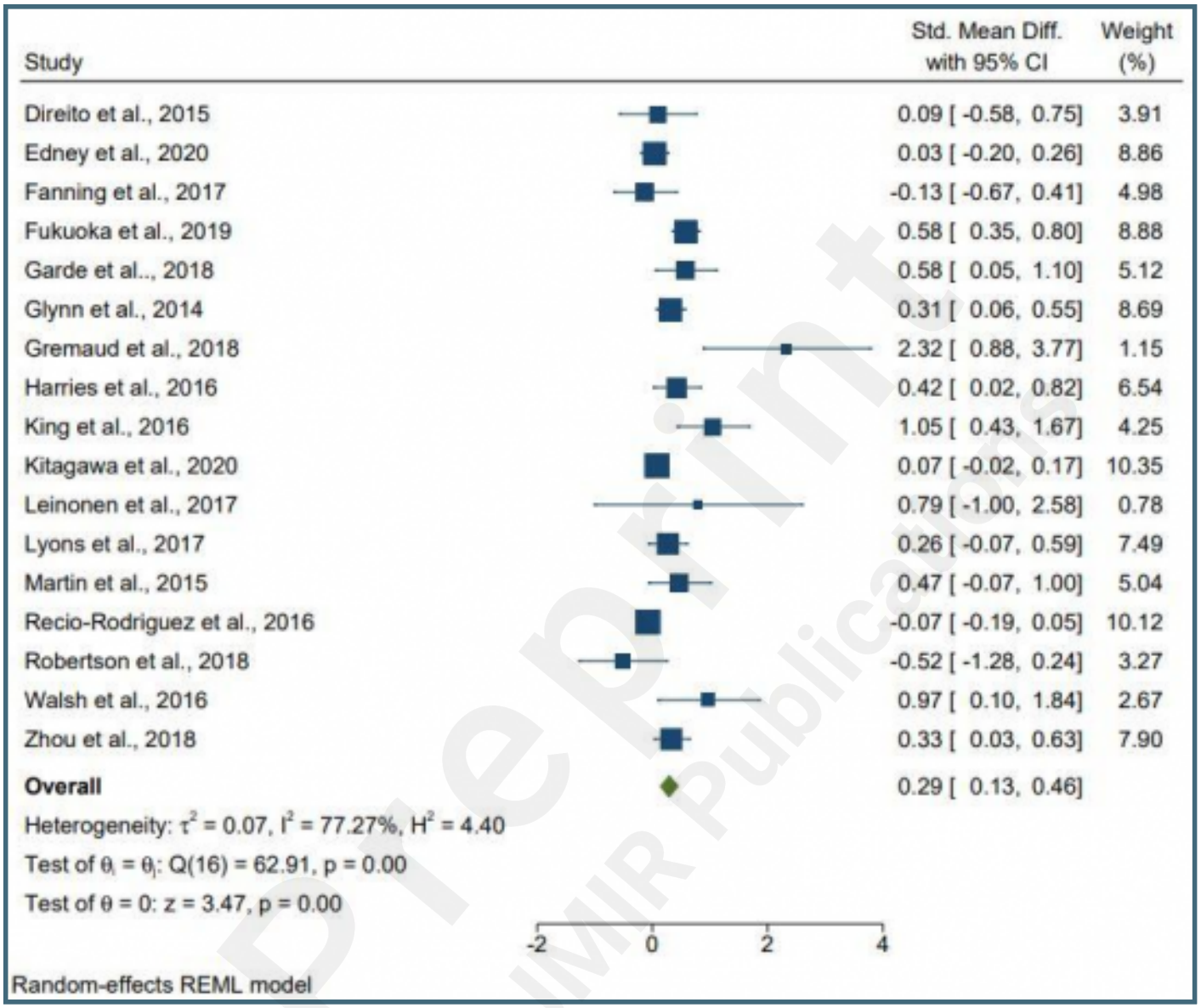
Untitled.
URL: http://asset.jmir.pub/assets/ba4fb2ec48b456084bd12e9fadd747e2.docx

Figures

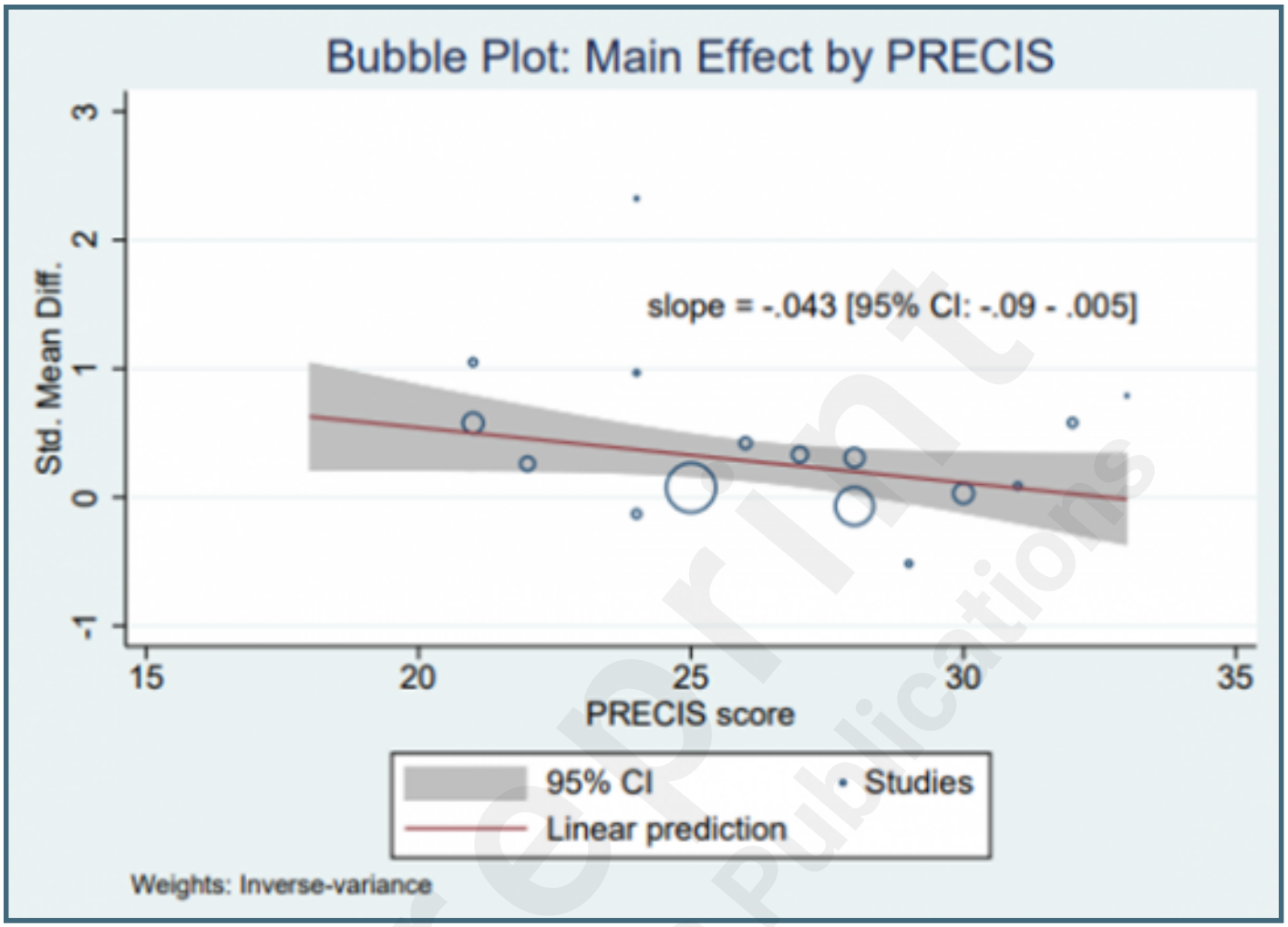
Flowchart of study selection.



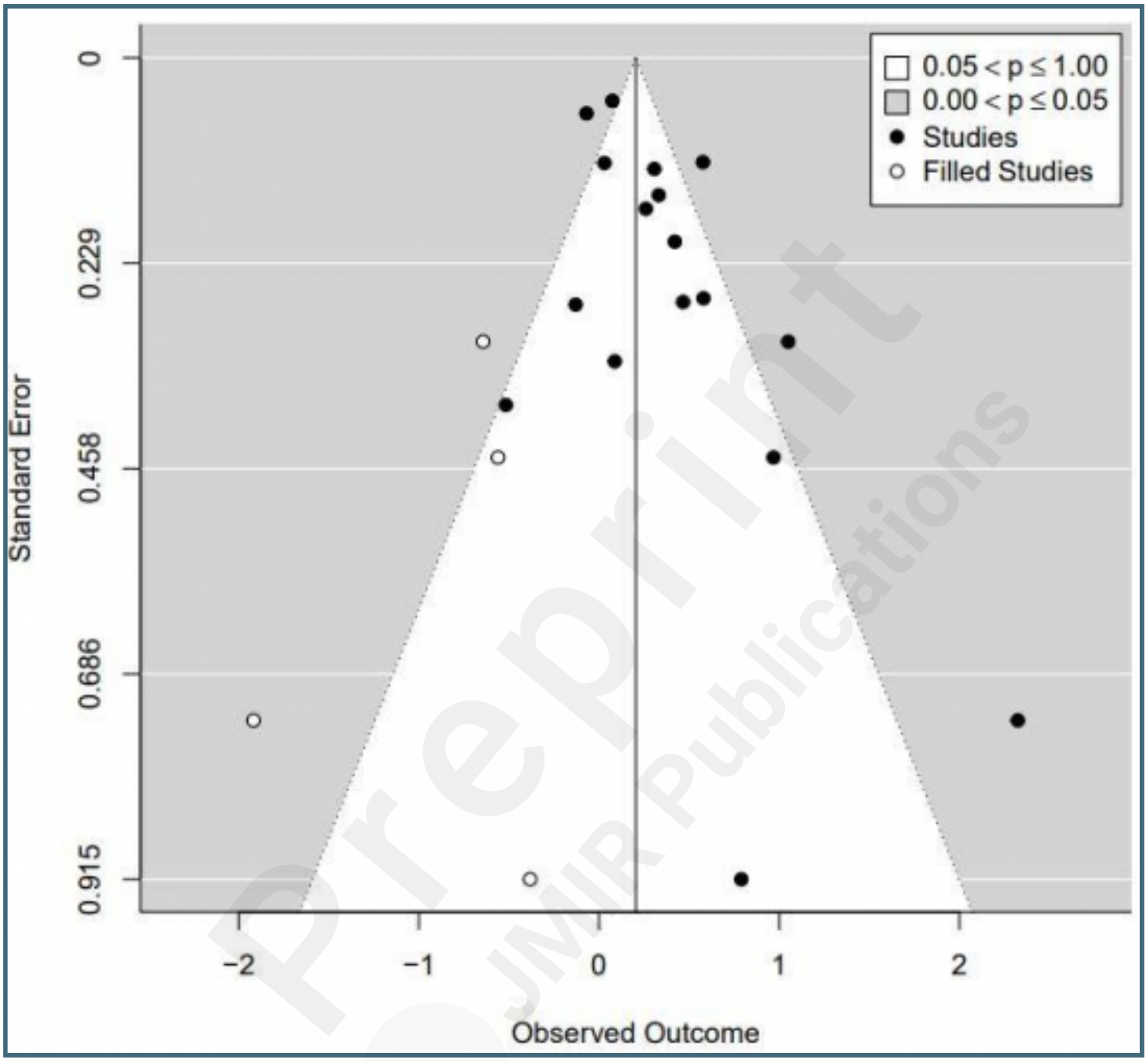
Forest plot of standardized treatment effects on physical activity with studies weighted by the inverse of the standard error of the estimated treatment effect.



Bubble plot of standardized treatment effect on PRECIS-2 score.



Trim-and-fill funnel plot for included studies in this meta-analysis.



Multimedia Appendixes

Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (doc).
URL: http://asset.jmir.pub/assets/4aab510c9891e1ee598e84efdc99e4c0.doc

Search Strategy for all Electronic Databases (pdf).
URL: http://asset.jmir.pub/assets/f57957c47f37867bf6a40d7b99360676.pdf

Data Extraction Form (General Study Characteristics) (pdf).
URL: http://asset.jmir.pub/assets/2680acc99205fcdc042db46790607740.pdf

Combined Coding Sheet (Adapted RE-AIM and PRECIS-2) (pdf).
URL: http://asset.jmir.pub/assets/f6fa4b0da912a30d32d058ee2ede5365.pdf

RE-AIM and PRECIS-2 scoring (pdf).
URL: http://asset.jmir.pub/assets/25c8490fd6f7a3bfc2ce83e2661e1249.pdf

Extracted study characteristics (doc).
URL: http://asset.jmir.pub/assets/9c92a855d55c514ded46ab08b5fd6e79.docx

Additional meta-analysis results (pdf).
URL: http://asset.jmir.pub/assets/4e367e7d8bd4535fdf23c33850436e5f.xlsx