Do physical activity interventions combining self-monitoring with other components provide an additional benefit compared with self-monitoring alone? A systematic review and meta-analysis

Tomas Vetrovsky,1,4 Agnieszka Borowiec,2 Roman Jiřík1, Charlotte Wahlich,3 Witold Śmigielski4, Michal Steffl1, James J Tufano1, Wojciech Drygas2,5 Petr Stastny1, Tess Harris3, Łukasz Małeķ2

ABSTRACT

Objective To determine the net effect of different physical activity intervention components on step counts in addition to self-monitoring.

Design A systematic review with meta-analysis and meta-regression.

Data sources Five databases (PubMed, Scopus, Web of Science, ProQuest and Discus) were searched from inception to May 2022. The database search was complemented with backward and forward citation searches and search of the references from relevant systematic reviews.

Eligibility criteria Randomised controlled trials comparing an intervention using self-monitoring (active control arm) with an intervention comprising the same treatment PLUS any additional component (intervention arm).

Data extraction and synthesis The effect measures were mean differences in daily step count. Meta-analyses were performed using random-effects models, and effect moderators were explored using univariate and multivariate meta-regression models.

Results Eighty-five studies with 12 057 participants were identified, with 75 studies included in the meta-analysis at postintervention and 24 at follow-up. At postintervention, the mean difference between the intervention and active control arms was 926 steps/day (95% CI 210 to 615). At follow-up, the mean difference was 413 steps/day (95% CI 210 to 615). Interventions with a prescribed goal and involving human counselling, particularly via phone/video calls, were associated with a greater mean difference in the daily step count than interventions with added print materials, websites, smartphone apps or incentives.

Conclusion Physical activity interventions that combine self-monitoring with other components provide an additional modest yet sustained increase in step count compared with self-monitoring alone. Some forms of counselling, particularly remote phone/video counselling, outperformed other intervention components, such as websites and smartphone apps.

WHAT IS ALREADY KNOWN

⇒ Self-monitoring of physical activity behaviour is a cornerstone of many complex physical activity interventions.
⇒ Self-monitoring using smartphone applications and activity trackers is effective in increasing physical activity levels.

WHAT ARE THE NEW FINDINGS

⇒ Additional intervention components further bolster the effect of self-monitoring.
⇒ Prescribed goal and human counselling are particularly effective in increasing physical activity above and beyond self-monitoring.
⇒ Remote phone/video counselling is a potentially highly effective and convenient component of physical activity interventions.

BACKGROUND

Given the high prevalence of physical inactivity and its associated public health burden,4 it is not surprising that researchers continue to explore which interventions can effectively and sustainably increase physical activity (PA).2 Such interventions are usually complex3 and consist of multiple components (eg, counselling, incentives, text messages and activity monitors).416 However, these complex interventions are typically assessed in parallel randomised controlled trials (RCTs) comparing an intervention arm to a usual care control arm. Consequently, these studies cannot determine which specific components actually contribute to the overall intervention effect and which components may have no effect or even a deleterious effect.11 Despite implementing alternative study designs such as factorial RCTs12 and microrandomised trials,13 the contribution of individual intervention components is still poorly understood.8 14

One of the most frequent components employed in complex PA interventions is self-monitoring, a crucial element of health promotion and disease prevention. Social cognitive theory states the importance of self-regulation as a source of behaviour change with self-monitoring one of the three core components.15 Historically, intervention participants recorded their daily PA in diaries,16 but self-monitoring with pedometers, fitness trackers and smartphone apps (collectively called activity monitors) has now become a cornerstone of PA interventions.17–19 A succession of systematic reviews has demonstrated that self-monitoring using different

© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

activity monitors leads to substantial PA increases which can be maintained long term.9 10 16 19 20

Considering the positive impact of self-monitoring on PA levels, it would be worth questioning if the effect of various complex PA interventions is primarily caused by self-monitoring alone and if additional components further bolster the effects or have no effects. For example, a recent meta-regression of step-count monitoring interventions compared with usual care suggested that interventions that also included counselling or incentives were not better than simpler interventions without counselling or incentives.10 Thus, it seems pertinent to question the net effect of various intervention components above and beyond the self-monitoring effects, especially as some of these components (eg, in-person counselling) are often resource-intensive compared with self-monitoring.21 22

Ultimately, the net effect of additional components (ie, after subtracting the effect of self-monitoring) can only be isolated in RCTs comparing the complex intervention including self-monitoring against self-monitoring alone (an ‘active control’ arm) rather than against a usual care control arm. Thus, this study’s aim was to conduct a systematic review and meta-analysis of these RCTs to determine whether complex PA interventions that combine self-monitoring using activity monitors with other intervention components provide an additional benefit to self-monitoring alone.

METHODS

This review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 statement.23 The protocol of this review can be found in the supplemental file online.

Eligibility criteria

This review included RCTs that compared an intervention using self-monitoring with an activity monitor to increase PA (active control arm) to an intervention that comprised precisely the same treatment as the active control PLUS any additional component intended to increase PA above the levels achieved in the active control condition (intervention arm).

The term ‘component’ is commonly used in the literature on complex interventions; however, it has no conventional definition.3 24 For the purpose of this study, we have introduced a working definition of a ‘component’ as a self-contained tool or channel that requires additional effort and cost to be developed, delivered and maintained. Examples of these components typically involve tools such as financial incentives or smartphone apps and channels such as in-person counselling or text messages.

Our use of the terms ‘active control’ and ‘intervention’ does not necessarily align with other authors’ labels. For example, studies often compared three arms: (1) usual care control, (2) pedometer-only intervention and (3) pedometer-plus-counselling intervention. In this specific case, the first arm would not be included in our review, the second arm would be labelled as ‘active control’, and the third arm as ‘intervention’. Furthermore, using the term ‘active control’ does not imply that this arm did not receive anything else other than the self-monitoring device. Active control arms often received print materials, education, etc, but they could still be eligible provided that the intervention arm received everything the active control arm received, plus some additional ‘intervention’ component(s). Studies where the active control arm received anything that was not also contained in the intervention arm were excluded, even when study authors labelled it as ‘active control’.
Studies comparing different intensities of the same component (e.g., low vs greater frequency of counselling, or sum of incentives), different types of the same component (e.g., individual vs team-based incentives), or additional behaviour change techniques within the same component (e.g., the addition of social comparison feature to an app) were ineligible. Studies were also excluded when active control participants were specifically instructed not to increase their PA despite receiving the activity monitor, or when they received a blinded monitor (for PA recording, but not for self-monitoring). Studies of dietary interventions where PA self-monitoring was also performed were included, provided that the intervention arm received an additional component with the aim to increase PA and not just to improve diet.

Studies in adults aged 18 or over, both healthy subjects and patients with specific diseases and conditions, were eligible. Both individually and cluster randomised trials were eligible. Finally, only studies that reported objectively assessed PA outcomes (e.g., step count, minutes of moderate-to-vigorous PA (MVPA)) were eligible.

Information sources, search strategy and selection process

An initial search of five databases (PubMed, Scopus, Web of Science, ProQuest, Discus) was performed in July 2020 and updated in May 2022. Search strings combined terms related to the domain being studied (e.g., “physical activity” or “steps”) AND intervention of interest (e.g., “monitor” or “pedometer”) AND terms signalling the presence of the active control arm (e.g., “active control” or “three arms”). Searches were limited to articles in English published in peer-reviewed journals but did not include any date limits. Details on the search strategy are presented in a supplemental file online.

Records were uploaded to EndNote V.X9, where duplicates were removed. Title and abstract and full-paper screening were conducted independently by two reviewers (ŁM and RJ) and discrepancies were resolved through discussion and consultation with the wider author team (TV, TH and PS).

In addition, backward and forward citation searches for the included articles and a search of the reference lists of the relevant systematic reviews were performed. The resulting records were uploaded to EndNote V.X9, deduplicated and screened for RCTs with an RCT classifier (https://robotsearch.vortext.systems) using the balanced machine learning model. Following the automatic screening, the remaining records were screened manually using the same strategy as with the database search.

Data collection process and data items

Data from included studies were extracted using a prepiloted Excel spreadsheet. Sample characteristics were jointly extracted by PS and RJ. A description of the active control and intervention conditions were first jointly extracted by LM, AB and WS. Using these excerpts, we (TV, LM and MS) developed a coding scheme by grouping the components into common-language categories such as ‘text messages’, ‘human counselling’ or ‘website’. For example, the ‘text messages’ category included traditional mobile phone SMSes but also tweets and push notifications. The ‘human counselling’ category was further divided into subcategories according to the mode of contact (e.g., face to face, email, and phone and video calls). The interventions were then coded for the presence of these components jointly by LM, AB, WD and WS., and independently in duplicate by TV. The discrepancies were resolved by consensus. In addition, whether
the participants received a specific goal for PA (e.g., 10,000 steps/day) was also coded.

PA outcomes were extracted by TV. Preferably, change scores from baseline to postintervention were extracted. When not available, preintervention and postintervention or postintervention only values were extracted. Occasionally, when authors used models to adjust for baseline measures of an outcome to reduce risk of bias due to high attrition rate, or to adjust for clustering

Figure 3  Mean difference in daily step count between the intervention and active control arms at postintervention (75 studies, 78 contrasts).
in cluster RCTs, estimates of effect were extracted directly. In addition, PA outcomes at the longest follow-up were extracted using the same strategy as for the outcomes at postintervention.

**Risk of bias assessment**

The risk of bias for each study was assessed by CW and independently assessed again by LM, WS, RJ or TV; disagreements were resolved through discussions with TH. The Cochrane Risk of Bias tool was used to assess the risk of bias (low, unclear or high risk) in each study for the following six domains: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data and selective reporting.

**Effect measures and synthesis methods**

The effect measures were the mean differences (MDs) in the daily step count between active control arms and intervention arms. The effect measures of cluster randomised trials were adjusted for clustering effects as recommended by the Cochrane Handbook. Meta-analyses were performed using random-effects models with a restricted maximum-likelihood estimator. Meta-analyses findings were presented using forest plots. The presence of publication bias was assessed using visual inspection of a funnel plot and random-effects version of Egger’s regression test. To check the robustness of the primary analysis, sensitivity analyses were carried out by excluding studies with active control arms (A) not having a set goal, (B) receiving an additional PA intervention component beyond activity monitor and a set goal, (C) receiving a non-PA intervention and (D) studies with small sample size (<40 participants).

The existence of heterogeneity was assessed using the I² statistic, which describes the percentage of total variation across studies that is due to heterogeneity rather than chance. The I² values were interpreted according to the Cochrane Handbook as not important (0%–40%), moderate (30%–60%), substantial (50%–90%) and considerable (75%–100%) heterogeneity. In case of substantial to considerable heterogeneity, effect moderators were explored using a series of univariate and multivariate meta-regression models. Potential moderators involved population characteristics (age, sex, body mass index (BMI), sedentariness, health status and healthcare setting) and intervention characteristics (intervention duration, presence of a set goal, intervention components including diary, print materials, website, smartphone app, text messages, incentives and human counselling). The amount of heterogeneity explained by the moderators included in the meta-regression models was expressed using R² statistics.

The significance level for all statistical tests was set at a p<0.05. All analyses were performed using the metafor package (V.2.0–0) in R V.3.4.4 (The R Foundation for Statistical Computing, Vienna, Austria).

**Differences from published protocol**

The final protocol of this review and meta-analysis slightly deviated from the originally registered protocol (online supplemental file 1, PROSPERO CRD42020199482) in the following aspects: (1) the EMBASE database was not searched because we lost access to it. However, Web of Science and ProQuest databases were searched instead of EMBASE; (2) As nearly all the studies reported daily step count as the PA outcome, we did not perform meta-analyses with minutes of MVPA and other outcomes as originally intended. Only 19 studies reported minutes of MVPA, and 15 of these also reported step count; (3) Consequently, not mixing step count and minutes of MVPA in one meta-analysis allowed us to use an absolute MD in daily steps instead of a standardised MD (Hedge’s g) as the measure of effect; and (4) The moderator analysis was performed using the meta-regression instead of subgroup analysis.

**RESULTS**

**Study selection**

We identified 57 eligible reports via the database search, 24 reports via backward and forward citation searches, and four reports by searching reference lists of relevant reviews. Altogether, 85 studies were included in the systematic review (figure 1).

Of those 85 studies, 75 studies provided sufficient details on the primary outcome at postintervention and were included in the primary meta-analysis. In most studies, multiple relevant arms were aggregated; however, in three studies, we chose to analyse the arms separately as they differed in factors explored in meta-regression (eg, in-person vs video coaching). In these three studies, we split the comparator group between the two arms to avoid double-counting it. Consequently, 78 contrasts were analysed in the primary meta-analysis at the end of the intervention.

Twenty-three studies also provided outcome data at follow-up. In addition, one study reported follow-up data in a separate report. In total, 24 studies and 25 contrasts were analysed in the meta-analysis at follow-up.

**Study characteristics**

Full details of the 85 studies included in the review are shown in the online supplemental file 3. The studies were published...
between 2007 and 2022. The majority of the studies were conducted in the USA (n=52), followed by the UK (n=10), Canada (n=3), Australia, Belgium, Japan, Singapore, Sweden (n=2 per country), Czechia, Denmark, Indonesia, Malaysia, the Netherlands, Pakistan, Poland, South Korea, Switzerland and Taiwan (n=1 per country). Thirty-five studies were conducted in healthcare settings, 18 at the university or college, and the remaining 32 studies were conducted in community or mixed settings. Three studies were cluster randomised trials.\(^{38} 91 117\)

The studies compared two (n=47), three (n=27), four (n=10) and five (n=1) arms. However, only 10 studies had 3 or more arms relevant to, and included in, the review. The number of study participants (counting only those from the relevant controls) ranged from 9 to 906 (median 80). In total, 12 057 participants were included in the review (9300 in the primary meta-analysis); of those, 60% were female. Participants’ mean age ranged from 20 to 76 (median 50), and their mean BMI ranged from 22 to 47 (median 30). Thirty-two studies recruited adults with chronic conditions (type 2 diabetes, overweight or obesity, hypertension, ischaemic heart disease, chronic obstructive pulmonary disease, breast cancer or Parkinson’s disease).

Intervention duration ranged from 1 week to 48 months (median 12 weeks); only 26 studies were longer than 3 months. Length of follow-up beyond the intervention ranged from 4 to 48 weeks (median 12 weeks). The types of self-monitoring devices used as the active control included pedometers (n=36), represented by predominantly Yamax (n=19) and Omron (n=13) brands; wearable fitness trackers (n=39) dominated by various models of Fitbit (n=27); and smartphone apps (n=8). In 55 studies, the active control arms received a specific goal of increasing PA. In 53 studies, the active control arms received at least one additional component intended to increase PA beyond self-monitoring: print materials (n=23); instructions to keep a step count diary (n=31); education and/or counselling (n=38). In 14 studies, the active control arms were exposed to a dietary (n=13) or sleep and stress (n=1) intervention. The contrasts between intervention and active control arms included addition of a set goal (n=27), print materials (n=12), step count diary (n=7), incentives (n=19), website (n=19), smartphone app (n=8), text messages (n=15), structured exercise (n=5) or human counselling (n=36). Counselling was provided in-person (n=10), as group counselling (n=7), via emails (n=9) or via phone or video calls (n=15).

### Risk of bias

Risk of bias judgements are presented in figure 2 with supporting quotes and justifications provided in the online supplemental file 4. Random sequence generation and reporting bias varied a lot between studies (54 low, 14 unclear and 17 high risk of bias). Allocation concealment was generally poorly reported, with a large number of studies being assessed as having an ‘unclear’ risk of bias for this domain (34 low, 49 unclear and 2 high risk of bias). Given the unavoidable risks associated with trials of behavioural interventions, which makes blinding unlikely, the majority of studies were judged to be at high risk of performance bias (7 low, 4 unclear and 74 high risk of bias). Outcome assessment (detection bias) was generally rated as low, due to the objective PA measures used (78 low and 7 high risk of bias). Most studies were judged to be at low risk of attrition bias; studies with high attrition rates (>20%),\(^{120}\) unclear reasons for drop-outs and disproportional incomplete outcome data across trial arms were deemed to be at high risk (62 low, 11 unclear and 12 high risk of bias). Finally, the risk of selective reporting of the outcome (reporting bias) varied between studies (53 low, 9 unclear and 23 high risk of bias).

### Effects at postintervention

At postintervention, the MD in daily step count between the intervention and active control arms (number of contrasts included in the analysis k=78) was 926 steps (95% CI 651 to 1201; I\(^2\) 87%) (figure 3). The visual inspection of the funnel plot (online supplemental file 5) complemented Egger’s regression test for funnel plot asymmetry (p=0.498) and indicated that publication bias was unlikely to have influenced the results.

Sensitivity analyses showed little change to the pooled effect sizes when excluding studies with active control arms receiving an additional PA intervention component beyond activity monitor and a set goal (k=32; MD 1023, 95% CI 772 to 1275; I\(^2\) 52%), studies with active control arms receiving a non-PA intervention (k=66; MD 915, 95% CI 613 to 1217; I\(^2\) 87%), studies with intervention duration <4 weeks (k=74; MD 883, 95% CI 599 to 1166; I\(^2\) 88%), and studies with sample size<40 (k=63; MD 824, 95% CI 523 to 1125; I\(^2\) 89%). However, when removing studies with active control arms not having a set goal, the MD between the intervention and active control arms substantially decreased (k=48; MD 710, 95% CI 459 to 961; I\(^2\) 78%). Excluding studies judged as low risk of bias in less than three of the six domains did not affect the pooled effect size (k=61, MD 912, 95% CI 587 to 1237; I\(^2\) 89%). However,
applying more stringent criteria and excluding studies with low risk of bias in less than four domains resulted in a decrease in the pooled effect size (k=41, MD 776, 95% CI 570 to 982, I² 63%).

Moderator analysis
Of the potential effect moderators related to the study population, neither age, percentage of females, BMI, being sedentary, suffering from chronic conditions, nor being recruited in a healthcare setting had a significant effect on daily step count in the univariate meta-regression models (table 1).

In the full multivariate meta-regression model that included intervention characteristics, only the intervention duration and addition of human counselling demonstrated significant moderating effects on daily step count (model 1, table 2). Starting with this model, we removed the non-significant variables one by one, starting with the one with the highest p-value, and the final model demonstrated a significant positive effect of presence of a set goal and additional counselling and a significant negative effect of longer intervention duration on daily step count (model 2, table 2). To further explore the effect of different types of human counselling, we replaced the ‘counselling’ variable in this model with dummy-coded binary variables representing individual counselling types. Of those, the addition of phone/video counselling had the largest effect on daily step count, followed by group counselling which, however, did not fall within the predefined statistical significance level. The addition of email and in-person counselling had small and non-significant moderating effects (model 3, table 2).

Effects at follow-up
At follow-up, the MD in daily step count between intervention and active control arms (k=25) was 413 steps (95% CI 210 to 615; I² 29%) (figure 4). Visual inspection of the funnel plot (online supplemental file 3) complemented Egger’s regression test for funnel plot asymmetry (p=0.767) and indicated that publication bias was unlikely to have influenced the results.

DISCUSSION
Combining other intervention components with self-monitoring using activity monitors provided an additional benefit of approximately 1000 steps/day compared with self-monitoring alone when assessed postintervention. At follow-up postintervention, this benefit reduced to less than half, but remained statistically significant at various time points.

Our findings are in contrast to Chaudhry et al’s conclusion that additional counselling/incentives offer no further benefit over simple self-monitoring interventions. However, their meta-regression primarily searched for studies examining the effects of step count monitoring devices compared with usual care; hence, their conclusion was based on indirect comparison of studies with and without counselling/incentives. In contrast, our systematic review and meta-analysis directly compared interventions combining self-monitoring and counselling/incentives with self-monitoring alone.

However, the net effect of additional components in our review (approximately 1000 steps/day) was substantially smaller than the effects of many complex interventions that also included self-monitoring, commonly ranging between 1500 and 2500 steps/day. Taken together with the previous evidence of the effect of self-monitoring alone (500–2000 steps/day), we can conclude that complex PA interventions owe part of their benefits to simple self-monitoring and their net effect is actually smaller than reported.

This conclusion is aptly illustrated by PACE-UP, the largest PA intervention trial to date. This 3-month complex intervention comprising three practice nurse consultations and pedometer self-monitoring increased PA by 1172 steps/day compared with usual care. A third active control arm, in which participants received a pedometer with a set step count goal, delivered by post without any counselling or support, increased their PA by 692 steps/day compared with usual care controls. Consequently, the nurse consultation net effect was only 480 steps/day, less than half of the intervention effect when compared with usual care. Thus, we argue that all studies exploring complex PA interventions that combine self-monitoring with other components should compare their intervention against an active control...
arm of self-monitoring with a set goal to isolate the additional components’ net effects, particularly as these additional components are often costly and resource-intensive.

The active control arm’s importance is further emphasised when addressing the maintenance of the intervention effect. At the 9-month follow-up of PACE-UP, the MDs between usual care and the (1) nurse counselling intervention and the (2) self-monitoring alone (active control group) were 677 steps/day and 642 steps/day, respectively. The nurse counselling net effect completely disappeared by the 9-month follow-up, while the pedometer-based self-monitoring effect was maintained in the same period and up to 3 years. Unlike PACE-UP, our meta-analysis demonstrated that the net effect of additional intervention components could be maintained for a certain period of time (median 3 months) if only reduced by half. Still, this finding does not weaken but rather supports our call for designing studies with active control arms.

Moderator analysis

Only intervention duration, the presence of a set goal, and the addition of human counselling (especially via phone/video calls) demonstrated significant moderating effects in the meta-regression models. However, none of these models explained more than 18% of the heterogeneity (as measured by $R^2$ statistics), which suggests that the between-study variability in effect sizes is being driven by other sources, which were not captured in the models.

Intervention duration had a negative effect on increases in step count: each additional week decreased the step count on average by approximately 10 steps/day. Even though this finding may seem intuitive as participant engagement can decrease over time, most meta-analyses did not detect any effect of intervention duration, and some found that longer interventions had a greater effect on PA. Thus, the impact of intervention duration on PA outcomes remains to be elucidated.

Our finding that setting a specific goal was associated with an increase of approximately 600 steps/day supports Bravata et al.’s 2007 finding that ‘having a step goal is the key predictor of increased PA’, as well as numerous subsequent meta-analyses. Indeed, given that provision of a set goal is so powerful, does not require any additional resources, and is closely interconnected with self-monitoring, we argue that a set goal should be an integral part of active control arms together with self-monitoring.

Human counselling was the only intervention component with a significant moderating effect with phone/video counselling having the greatest effect. We are unaware of any review directly comparing phone/video with in-person or other types of human PA counselling. However, several remote feedback reviews indicated that phone/video calls are indeed a promising way of PA counselling. Thus, the use of phone/video counselling in PA interventions should be further explored as it offers a potentially highly effective and flexible tool, especially useful during the COVID-19 pandemic.

It is important to realise that not having a significant moderating effect, as was the case for financial incentives, text messages or in-person counselling, does not mean that the addition of these components had no impact on PA. The regression coefficients and their p-values reported in table 2 represent the moderating effect of the intervention component on the MD between the intervention and active control arms, not the total effect of interventions including this component. Therefore, the results of our moderator analyses do not contradict previous meta-analyses that demonstrated a significant effect of text messages, financial incentives, and smartphone apps on PA levels.

Studies exploring the financial incentives commonly use some form of objective self-monitoring and a set goal (frequently expressed in daily step count), as the incentives are usually contingent on meeting this goal. Similarly, most smartphone apps aimed to increase PA also include self-monitoring, either using internal smartphone accelerometers or accompanying wearables. However, unlike studies of financial incentives that almost always use self-monitoring alone as the active control arm, studies of smartphone apps rarely do. Therefore, the previous meta-analyses of smartphone apps could not distinguish whether the effect of the apps resulted mostly from self-monitoring alone or other additional features of the app. Thus, in line with our previous reasoning, we argue that future RCTs of smartphone apps should always consider including an active control arm (using an app with stripped features but still allowing the self-monitoring) as is already common in RCTs of financial incentives.

Strength and limitations

This is the first meta-analysis capable of isolating the additional benefit of various intervention components above and beyond simple self-monitoring. The high number of included studies allowed for robust detection of the net effect of these components at both the postintervention and follow-up time points. Additionally, we conducted a series of meta-regression models sufficiently powered to demonstrate the moderating effect of intervention duration, set goal and additional human counselling, especially phone/video counselling, supporting the findings of previous meta-analyses.

A potential limitation is in our identification of all eligible studies. As there is no easy way to detect studies comparing a PA intervention against an active control arm (only 11 out of 85 included studies used the term ‘active control’), our search relied on proxy terms, such as ‘three arms’, ‘with and without’ or ‘self-monitoring alone’. Thus, we could not be sure that our initial database search identified all eligible studies. We, therefore, applied a rigorous backward and forward citation search and also searched reference lists of six relevant systematic reviews to strengthen the database search. Even if we cannot ensure that we identified all eligible studies, the high number of studies included in the meta-analysis (n=75) makes it unlikely that a few more studies would have considerably affected our findings, as suggested by the series of sensitivity analyses.

A further limitation is the lack of clarity around the term ‘component’. Our review identified studies employing conceptually very different approaches (eg, text messages, counselling, and financial incentives). While the scheme developed for coding the components is well fit for the purpose of this study, it has not been rigorously validated and limits comparability with other studies. Furthermore, some codes are very broad and involve components with various levels of sophistication (eg, the ‘website’ category spans simple educational websites as well as advanced gamified online programmes). Thus, future research would benefit from the development of a standardised taxonomy of intervention components, analogous to Michie’s taxonomy of behaviour change. While it was not the objective of this study, a refined coding of the intervention components using Michie’s taxonomy could also provide us with a more robust insight into precisely which behaviour change techniques act as moderators. It might also be interesting to explore whether interventions are
based on theories (ie, the social cognitive theory) and explore the associated mechanisms that have been addressed with particular intervention characteristics.

Next, our findings that resulted from the meta-regression models need to be interpreted with caution due to the observational nature of meta-regression. Indeed, meta-regression can be subject to many pitfalls, and our conclusions on the differential effect of various components must be considered preliminary and need to be confirmed in future RCTs isolating the effects of individual components using active control arms or a factorial design.

In addition, the active control arms of many of the studies included in the meta-analysis also included some intervention components in addition to self-monitoring. Even though these components were usually minimal (eg, a leaflet with general information on the benefits of PA, one short educational session), they might have reduced the estimated effect of the additional component in intervention arms. However, the results of the sensitivity analysis excluding studies with active control arms receiving an additional PA intervention component beyond activity monitor and a set goal showed only a little increase in the pooled effect sizes when compared with the primary analysis (1023 vs 926 steps/day).

Finally, the findings of this meta-analysis can be inflated by low-quality studies. In fact, only 39 of 75 studies included in the meta-analysis at post-intervention were judged as having a low risk of bias in at least four out of six domains. The sensitivity analysis limited to these high-quality studies showed a considerable decrease in the pooled effect of additional components by approximately 150 steps/day when compared with the meta-analysis of all studies. Thus, the importance of using self-monitoring with a set goal as active control groups seems to be all the more pertinent.

**Implications for practice, policy and future research**

Findings from this review and other literature suggest that researchers should consider including self-monitoring using a simple activity monitor (or a smartphone app) together with a set goal (eg, adding 3000 steps/day to usual PA levels) as a fundamental component of all complex PA interventions. Furthermore, researchers assessing complex PA interventions in RCTs should compare the intervention approach against an active control arm consisting of self-monitoring and a set goal to isolate the net effects of additional intervention components. Along the same lines, policy-makers and practitioners should prioritise simple self-monitoring interventions (eg, pedometer-based or via smartphone apps) unless the effects of additional resource-intensive components above and beyond self-monitoring alone are rigorously demonstrated. Specifically, advanced smartphone apps should be researched to determine whether they are a cost-effective means of increasing PA, as their effects might predominantly result from their self-monitoring capacities.

Finally, this review indicated that some forms of human counselling, particularly remote phone/video counselling, can be potentially very effective components of PA interventions. Given their lower costs compared with in-person counselling, and considering their convenience and suitability in times of pandemics and other scenarios that do not allow for face-to-face contact, remote counselling (complemented with self-monitoring and a set goal) should be further explored as a promising way of combatting yet another pandemic—that of physical inactivity.

**CONCLUSIONS**

Complex PA interventions that combine self-monitoring using activity monitors with other intervention components provide an additional benefit above and beyond self-monitoring alone. Some forms of human counselling, particularly remote phone/video counselling, outperformed other intervention components such as websites and smartphone apps.

**REFERENCES**


Cherubini M, Willebois-Zuliga G, Boldi M O. The unexpected downside of paying or sending messages to people to make them walk: comparing tangible rewards and motivational messages to improve physical activity. *Acm T Comput Hum Int* 2020;27:1–44.


to enhance social incentives to increase physical activity among families: a controlled trial using different types of feedback on employees’ physical activity. J Telemed Telecare 2019;25:151–7.


Pate RL, Benjamini EJ, Volpe KG, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the be fit randomized clinical trial. JAMA Intern Med 2017;177:1586.


