Citation for published version:
McEwan, D, Rhodes, RE & Beauchamp, MR 2020, 'What Happens When the Party is Over? Sustaining Physical Activity Behaviors after Intervention Cessation', *Behavioral Medicine*, pp. 1-9.
https://doi.org/10.1080/08964289.2020.1750335

DOI:
10.1080/08964289.2020.1750335

Publication date:
2020

Document Version
Peer reviewed version

Title: What Happens when the Party is Over?: Sustaining Physical Activity Behaviors after Intervention Cessation

Authors & Degrees: Desmond McEwan (PhD)[1*], Ryan E. Rhodes (PhD)[2], Mark R. Beauchamp (PhD)[3]

Affiliations:
[1]: University of Bath, Department for Health, Bath, United Kingdom
[2]: University of Victoria, School of Exercise Science, Physical & Health Education, Victoria, Canada
[3]: University of British Columbia, School of Kinesiology, Vancouver, Canada

Contact Information for Corresponding Author*:
Mailing Address:
University of Bath
Claverton Down
1W 4.108
Bath, Somerset, United Kingdom
BA2 7AY

Email Address:
d.a.mcewan@bath.ac.uk

Twitter:
@desmondmcewan

The Version of Record of this manuscript has been published and is available in

Behavioral Medicine

https://www.tandfonline.com/doi/full/10.1080/08964289.2020.1750335

Abstract

Although extensive research suggests that behavior change interventions can improve physical activity (PA) over the course of an intervention, the maintenance of these improvements beyond intervention termination is less clear. The purpose of this study was to determine, through meta-analysis, whether behavior change interventions produce sustained improvements in PA after interventions conclude. Studies were retrieved from a recent (2019) meta-analysis of 224 interventions. Studies that measured PA at baseline, post-intervention, and a follow-up timepoint were included in this updated review. We examined the effects of these interventions in terms of changes in PA from baseline to post-intervention, baseline to follow-up, and post-intervention to follow-up (relative to control groups). We also examined whether the inclusion of theory and behavior change techniques (BCTs) within interventions as well as the length of time between PA assessments moderated these effects. Thirty-nine interventions (17% of interventions from the previous review) from 31 studies were included in the meta-analysis. Significant improvements in PA were found from baseline to follow-up (d = 0.32, p < .001). In general, these effects resulted from significant increases in PA from baseline to post-intervention (d = 0.46, p < .001), followed by significant decreases from post-intervention to follow-up (d = -0.18, p = .010). Effect sizes did not vary between theory-based and no-stated-theory interventions. The positive effects from baseline to post-intervention and negative effects from post-intervention to follow-up were more pronounced as the length of time between assessments increased. In conclusion, behavior change interventions improve PA over the course of the intervention; however, these improvements are generally not sustained after the intervention concludes.

Keywords: behavior change; exercise; health; meta-analysis; theory

INTRODUCTION

Recent estimates show that approximately one out of three adults and four out of five adolescents do not partake in sufficient levels of health-enhancing physical activity (PA).[1] As a result, it is essential to understand how PA behavior can be improved.[2] In response to this need, a wide range of behavior change interventions targeting PA have been developed.[2] These interventions typically utilize one or more *behavior change techniques* (BCTs), which are described as the 'active ingredients' of an intervention—that is, the observable, replicable, and irreducible components that are designed to modify the processes that regulate behavior.[3] Sixteen clusters of BCTs—comprised of 93 lower-order strategies—have been identified, including: goals and planning; feedback and monitoring; social support; shaping knowledge; natural consequences; comparison of behavior; associations; repetition and substitution; comparison of outcomes; reward and threat; regulation; antecedents; identity; scheduled consequences; self-belief; and covert learning.[3] In some interventions, the delivery of these BCTs is guided by a theory of behavior change, which provides a framework for systematic implementation and evaluation.[4] In other interventions, those same BCTs may be used but their delivery is not guided by explicit use of psychological theory.

Somewhat encouragingly, several meta-analytic reviews have found positive and significant effects of behavior change interventions to improve PA.[5,6] However, in most reviews, these effects are delimited to PA measurements at baseline and post-intervention.[7] Thus, although the research evidence to date suggests behavioral interventions can produce significant improvements in PA over the course of the intervention, it is less clear whether these effects are maintained after the intervention concludes.[7,8,9] Moreover, among previous reviews of interventions that have examined PA changes over the longer-term, the bulk of evidence appears to suggest that PA behaviors return to baseline levels as the length of time between assessments increases.[8] For example, a meta-analysis by Murray et al. found that interventions produced significant improvements in PA at 6-9 months post-baseline (relative to controls) but these effects decreased in size at 9-15 months; beyond 15 months, the effects were small and inconsistent.[8] It should be noted, however, that this study did not examine physical activity behaviors after the cessation of the intervention (i.e., from post-intervention to a follow-up timepoint).[8] As such, it would seem particularly important to identify whether behavior change interventions not only enable the *uptake* of PA behavior during the course of a given intervention, but also (and especially) the *maintenance* of PA after an intervention has ended.

As a complement to the above question, it would appear informative to assess whether the use of theory and/or BCTs impact the post-intervention effects on PA. This could improve our understanding of the components that are necessary to include in an intervention in order to not only generate initial increases in PA but the preservation of these improvements over time. For example, certain BCTs might be useful in improving PA over the course of an intervention; however, if PA levels then return to baseline levels after an intervention has ended (i.e., at a follow-up timepoint), the veracity of these BCTs to truly change PA would appear to be rather limited. In addition, although a growing body of research seems to suggest that theory-based interventions are no more effective for improving PA over the course of an intervention,[2,6,10,11] it would be worth examining whether theory influences the *maintenance* of those PA increases after the intervention has ended.

The purpose of the present study was to conduct a meta-analysis examining the effectiveness of behavior change interventions in producing sustained improvements in PA (i.e., after the interventions ended). To do so, we reviewed papers included in a recent meta-analysis that examined whether interventions improved PA behavior (relative to control conditions) based on their use of theory and behavior change techniques.[6] As that meta-analysis included papers from 68 previous reviews, it provides a comprehensive collection of the existing published studies that have tested the effects of behavior change interventions on PA. Nonetheless, it should be noted that the effect sizes in that meta-analysis were based on changes in PA from baseline to post-intervention, which only provides evidence of the effectiveness of behavior change interventions in producing initial improvements in PA over the course of an intervention (i.e., the uptake of PA behavior). Therefore, we sought to re-analyse data from studies subsumed within that previous review[6] in order to determine whether the initial effects on PA were maintained following the conclusion of the intervention (i.e., at a follow-up time point). We also sought to examine whether the use of theory and/or specific BCTs moderated the effects of these interventions on PA across three timeframes: baseline to post-intervention, post-intervention to follow-up, and baseline to follow-up. In addition, we sought to assess whether the potential changes in PA across these timeframes were moderated by the length of time between measurements.

Based on the findings from previous reviews,[2,6,8,10,11] we hypothesized that significant improvements in PA from baseline to post-intervention would be shown (hypothesis #1a) and that these effects would be evident regardless of the length of time between assessments (hypothesis #1b), explicit use of theory to guide intervention delivery (hypothesis #1c), or specific type of BCT cluster incorporated in the intervention (hypothesis #1d). In light of the evidence suggesting that long-term improvements in PA are small and inconsistent from baseline to follow-up periods[8], we anticipated that similar findings would emerge in our review. Specifically, we hypothesized that PA would improve to a small extent from baseline to follow-up (hypothesis #2a), and that effects over this timeframe would be smaller as the length of time between assessments increased (hypothesis #2b). It was also hypothesized that those effects would not be moderated by the inclusion of theory (hypothesis #2c) or any particular BCT clusters (hypothesis #2d) within interventions. Finally, we hypothesized that there would be significant decreases in PA from post-intervention to follow-up (hypothesis

#3a) and that these negative effects would be larger as the length of time between assessments increased (hypothesis #3b). We did not expect the inclusion of theory (hypothesis #3c) or any specific type of BCT cluster (hypothesis #3d) to moderate the effects over this timeframe.

METHODS

All papers included in the previous review[6] were reviewed for inclusion in the current meta-analysis. Full details on the methods employed to obtain articles are noted in the original paper as well as in the pre-registered protocol on PROSPERO.[12] In brief, the effects of 224 controlled interventions from 171 studies were examined in that meta-analysis and compared in terms of their use of theory and BCTs. Studies were included in that review if they met the following inclusion criteria: (1) controlled experimental intervention; (2) intervention targets physical activity only (i.e., interventions targeting multiple health behaviors were excluded); (3) must have an outcome measure of physical activity that is assessed at both baseline and post-intervention; (4) published in a peer-reviewed journal (i.e., conference proceedings and dissertations were excluded); and (5) provide sufficient data to compute an effect size. The studies from that meta-analysis were reviewed and subsequently included in the current meta-analysis if—in addition to the above inclusion criteria—PA was also measured at a follow-up period (i.e., at some point after the behavior change intervention had ended). In other words, studies were included in this meta-analysis if effect sizes could be computed for changes in PA from baseline (i.e., pre-intervention) to post-intervention, post-intervention to follow-up, and baseline to follow-up.

A range of PA measures were used in the included studies (e.g., self-report questionnaires, pedometers). Where possible, standardized effect sizes for each study were computed based on means, standard deviations, and sample sizes in relation to the PA variable (e.g., steps per day, minutes of moderate-to-vigorous PA per week) for experimental and control conditions at each timepoint. If such statistics were missing, we used F-values, *t* statistics, and *p* values to compute an effect size.[16] Each study was given a relative weight based on its precision, which is determined by the study's variance, standard error, and confidence interval (i.e., more precise data is given a larger relative weight compared to less precise data).[16] These weighted effect sizes were combined and data were analysed for each timeframe as random-effects models using Comprehensive Meta-Analysis, Version 2 software.[13] Cohen's *d* was computed to represent the standardized effect of an intervention on PA—a value of 0.2 corresponds to a 'small' effect, 0.5 to a 'medium' effect, and 0.8 to a 'large' effect.[14] For studies that examined physical activity outcomes for multiple experimental conditions in comparison to the same control condition, potential unit-of-analysis errors were corrected by dividing the sample size of the control condition by the number of within-study comparisons.[15]

Heterogeneity was evaluated with Q-values which estimate the variability in the observed effect sizes across studies, as well as I^2 values which estimate as the ratio of the true heterogeneity to the total observed variation.[16] The level of heterogeneity can be interpreted with the following estimates of I^2: 0-40%=trivial; 30-60%=moderate; 50-90%=substantial; and 75-100%=considerable.[15] To assess the impact of each individual intervention and identify any outliers, sensitivity analyses were conducted by computing the overall effect size after an intervention is removed from the analysis.[16] Outliers were identified by considering each of the following.[16] First, an intervention was flagged as a potential outlier if it had a very large effect size and narrow confidence interval, thus producing an abnormally high Z-score (i.e., that is, compared to the Z-scores of other studies in the meta-analysis). Second, we considered whether the removal of an intervention (i.e., $k - 1$) from the analysis resulted in an effect size that differed substantially from the original effect size (i.e., *k*). Third, if heterogeneity (Q and I^2) is reduced upon removal of an intervention, this provides further evidence that it is an outlier and should be omitted from subsequent subgroup/moderator analyses, as it could potentially skew the results of those analyses. As with the original meta-analysis,[6] publication bias was considered using fail-safe N values, Duval and Tweedie's trim-and-fill statistics, and funnel plots.[16] Publication bias may be present if fail safe *n* values do not exceed 5*n* + 10 or if point estimates for Duval and Tweedie's trim-and-fill assessment differ substantially from the original effect size.[16,17]

The included studies were coded by two independent coders in order to examine three potential moderators: (a) length of time (in weeks) between assessments for each timeframe (i.e., from baseline to post-intervention, baseline to follow-up, and post-intervention to follow-up); (b) their use of theory ('theory-based' or 'no-stated-theory' intervention); and (c) the presence of the 16 BCT clusters from Michie et al.'s behavior change taxonomy ('yes' or no' for each of 'scheduled consequences', 'rewards and threat', 'repetition and substitution', 'antecedents', 'association', 'covert learning', 'natural consequences', 'feedback and monitoring', 'goals and planning', 'social support', 'comparison of behavior', 'self-belief', 'comparison of outcomes', 'identity', 'shaping knowledge', and 'regulation').[3,6,12] A meta-regression—using maximum likelihood estimation—was carried out to examine whether the continuous variable of the length of time (in weeks) between measurements moderated the effect sizes for each timeframe. Q-values and I-squared values are provided for each meta-regression model as a means of estimating whether this predictor (i.e., length of time between assessments) accounts for part of the between-study variability. Subgroup analyses were conducted to examine the potential moderating effects with regard to the categorical variables (i.e., 'yes' versus 'no') of theory use and presence of each BCT cluster. Specifically, interventions guided by a theory were pooled together and this pooled effect size was compared to the pooled effect size of interventions that were not guided by a theory. Similarly, for each of the 16 BCT clusters, we compared the pooled effect size of interventions that included the BCT cluster versus those that did not. A Q statistic was produced for these pooled comparisons; significant moderation is evident if $p < .05$ for this Q-value. Borenstein et al.[16] suggest that effect sizes derived from subgroups with five or fewer interventions can be inaccurate/misleading. Thus, although we provide the results of all subgroups for descriptive purposes (except for those based on a single intervention), the majority of our discussion herein is devoted to the results based on six or more interventions; effect sizes based on fewer than six interventions should be interpreted with caution.

RESULTS

Among the 171 studies from the original meta-analysis, 31 studies (18%) included a follow-up measure of PA. Six studies included multiple interventions, which resulted in k = 39 total comparisons (i.e., 17% of the 224 interventions from the original meta-analysis).[6] A list of studies is provided in supplementary material 1; effect sizes and accompanying statistics for each intervention, along with sensitivity analyses and forest plots are shown in supplementary material 2. Overall, there were significant, small-to-medium-sized improvements from baseline to follow-up ($d = 0.32$, $p < .001$). These effects appeared to result from significant, medium-sized improvements in PA from baseline to post-intervention ($d = 0.46$, $p < .001$), followed by significant, small-sized decreases from post-intervention to follow-up ($d = -0.18$, $p = .010$). Considerable heterogeneity was evident for each timeframe. Specifically, Q = 229.80 ($p <.001$) and $I^2 = 83.64$ from baseline to follow-up; Q = 580.24 ($p <.001$) and $I^2 = 93.45$ from baseline to post-intervention; and Q = 238.75 ($p <.001$) and $I^2 = 84.04$ from post-intervention to follow-up. Publication bias did not appear to be present, as indicated by (a) fail safe n values exceeding Rosenberg's critical value of $5n + 10$ (205) and (b) Duval and Tweedie's point estimates not differing from the original effect sizes (see supplementary material 3).[16,17]

One outlier was detected[18] (see supplementary material 2). Specifically, from baseline to follow-up, this study had a Z-score of 17.38; by comparison, all other studies had a Z-score between -0.03 to 3.59. Moreover, when this study was removed, the effect size was *d* = 0.28 (p < .001) from baseline to follow-up; the effect size when any other study was removed ranged from *d* = 0.31 to 0.32. Similarly, from baseline to post-intervention, this study had a Z-score of 26.91, whereas all other studies had a Z-score between -0.85 to 6.15. When this study was removed, the effect size for this timeframe was *d* = 0.39 (p < .001); for all other studies, the removal of the study resulted in an effect size between *d* = 0.43 to 0.47. From post-intervention to follow-up, this study had a Z-score of -14.34; in comparison, all other studies had a Z-score between -3.84 to 2.03. The effect size when this study was removed was *d* = -0.13 (p = .007) for this timeframe. For all other studies, the removal of a study resulted in an effect size between -0.16 to -0.19. Although still significant (in terms of Q-values) and substantial (in terms of I^2 values), heterogeneity across studies for each timeframe also decreased when this study was removed. As mentioned above, Q = 580.24 and I^2 = 93.45 from baseline to post-intervention when this study was included; these values reduced to Q = 131.43 (p <.001) and I^2 = 71.85 when it was excluded. From baseline to follow-up, Q = 229.80 (p <.001) and I^2 = 83.64 when this study was included; Q = 74.15 (p <.001) and I^2 = 50.10 when it was excluded. From post-intervention to follow-up, Q = 238.75 and I^2 = 84.04 when this study was included; Q = 94.10 (p <.001) and I^2 = 60.68 when it was excluded. Given the above findings, this study was excluded from the subsequent moderator analyses.

The length of time between assessments from baseline to post-intervention ranged from 2 to 26 weeks. The Q-value for this meta-regression examining the potential moderating effect of time between assessments was 43.58 (p = .212) while I^2 = 15.10; the slope for this predictor (time) was positive and significant (B = 0.018, p = .041). This suggests that length of time between assessments accounts for part of the between-study variability and that effect sizes were larger as the time between assessments increased. The length of time between assessments from baseline to follow-up ranged from 12 to 52 weeks. For this meta-regression

Q = 46.65 (p = .445) and I^2 = 1.40; the slope examining the effect of time was not significant (B < 0.000, p = .951). Hence, the effect size for this timeframe did not appear to be moderated by the length of time between assessments. Finally, the length of time between assessments from post-intervention to follow-up ranged from 6 to 48 weeks. For this meta-regression Q = 54.92 (p = .173) and I^2 = 16.24; the slope examining the effect of time was negative and significant (B = -0.010, p = .049). This suggests that the negative effects from post-intervention to follow-up were larger in size (in the negative direction) as the time between assessments increased.

The results of the categorical moderator analyses examining the potential effects of theory as well as BCT clusters are provided in Table 1. Significant improvements in PA were found from baseline to post-intervention among theory-based interventions (k = 28, d = 0.44) as well as no-stated-theory interventions (k = 10, d = 0.29), with no significant difference in effect sizes between these two pools of studies (Q = 1.12, p = .29). Similarly, there were significant improvements from baseline to follow-up for both theory-based interventions (d = 0.30) and no-stated-theory interventions (d = 0.28)—these effect sizes did not differ significantly (Q = 0.04, p = .85). From post-intervention to follow-up, there was a significant decrease in PA among theory-based interventions (d = -0.15) but no significant change among no-stated-theory interventions (d = -0.09); the difference between these two pools of interventions was not significant (Q = 0.23, p = .63).

With regard to the influence of BCTs, there were significant improvements among all interventions (i.e., those that included any BCT cluster) from baseline to follow-up (ds ≥ 0.28) (see Table 1). In the majority of interventions, this effect resulted from a significant improvement in PA from baseline to post-intervention, followed by a significant decrease from post-intervention to follow-up. However, there were two notable exceptions to these findings. Specifically, we found small, significant improvements in PA from baseline to post-intervention for ‘comparison of outcomes’ and ‘self-belief’; these improvements appeared to be maintained over time, as there were no significant changes in effect sizes from post-intervention to follow-up. Similar findings across each timeframe were also demonstrated for ‘association’, ‘antecedents’, and ‘identity’; however, the results for these three BCT clusters need to be interpreted cautiously as they were based on four, four, and three interventions, respectively.

As a supplement to the above analyses, Q-statistics were computed for each BCT cluster wherein the effect sizes of interventions that included a cluster were compared to those that did not include the BCT cluster. From pre- to post-intervention, larger effect sizes were shown for interventions that incorporated ‘goals and planning’ (Q = 4.42, $p = .036$), ‘natural consequences’ (Q = 6.63, $p = .010$), ‘comparison of behavior’ (Q = 6.43, $p = .011$), or ‘scheduled consequences’ (Q = 6.33, $p = .012$) relative to those studies that did not include the respective BCT within the intervention. From pre-intervention to follow-up, significantly larger effect sizes were noted in interventions that included ‘antecedents’ (Q = 3.91, $p = .048$) compared to those that did not incorporate this type of BCT. Finally, from post-intervention to follow-up, interventions that did not include ‘comparison of outcomes’ (Q = 3.88, $p = .049$) as well as those that did not include ‘self-belief’ (Q = 3.95, $p = .047$) showed greater decreases in PA compared to those studies that did include these BCT clusters (which, as mentioned above, did not demonstrate significant changes in PA over this timeframe).

DISCUSSION

The purpose of this review was to examine, through meta-analysis, the effects of behavior change interventions on PA from baseline to post-intervention, baseline to follow-up (i.e., a timepoint following the completion of the intervention), and post-intervention to follow-up. Overall, we found significant, medium-sized improvements in PA from baseline to post-intervention, followed by significant, small-sized decreases from post-intervention to follow-up—this ultimately resulted in significant, small-to-medium-sized increases in PA from baseline to follow-up. For each of these three timeframes, we sought to examine the potential impact of the length of time between measurements, as well as the inclusion of theory and BCT clusters within the intervention. The findings concerning each of these moderating variables, along with associated limitations and suggestions for future research, are discussed, in turn, below.

The length of time between PA measurements appeared to moderate the effect sizes that were found over the course of the intervention and following intervention cessation. Specifically, the positive, significant effects from baseline to post-intervention were larger as the length of time between assessments—which ranged from 2 to 26 weeks—increased. This might be due to the increased amount of contact time and total intervention components that participants receive over the course of longer interventions compared to shorter ones.[19] With regard to changes in PA from baseline to follow-up, there was no moderating effect of time between assessments (which ranged from 12 to 52 weeks). This null finding appears to align with the results from previous reviews. In particular, Murray et al.[8] found small, positive effects of interventions on PA and these effects were maintained up to 15 months (~65 weeks) following baseline. Finally, a significant moderating effect of time between assessments was demonstrated from post-intervention to follow-up (which ranged from 6 to 48 weeks). Specifically, the negative effect sizes in PA over this timeframe were more pronounced as the length of time between assessments increased. The aforementioned review by Murray et al.[8] found inconsistent effects of interventions beyond 15 months (post-baseline). As a complement to those findings, our results seem to suggest that the initial improvements obtained over the course of PA behavior change interventions dissipate as the time lapse between post-intervention and follow-up assessments increases. It should be noted that in the review by Murray et al.[8], interventions were pooled together into five intervals according to the time at which PA was measured (6-9, 9-15, 15-21, 21-24, >24 months post-baseline), regardless of the duration of the intervention itself. Hence, effect sizes computed up to 15 months could include interventions that were still ongoing at that timepoint as well as those that had already ended previously. In contrast, we specified “follow-up” as the measurement of PA after a period following the conclusion of an intervention. As such, our findings add to the results by Murray et al.[8] by demonstrating that it is not merely the time since the initial baseline timepoint that plays a role in predicting PA levels, but also (and more specifically) the time since the cessation of the intervention.

With regard to the impact of theory, positive medium-to-large effects on PA from baseline to post-intervention were demonstrated for interventions that were explicitly guided by a theory as well as those that were not. Similar to the findings from the meta-analysis by McEwan et al.[6], the effect size appeared slightly larger for theory-based interventions than for no-stated-theory interventions, but this difference was not significant. Interestingly, although there were significant decreases in PA from post-intervention to follow-up among theory-based interventions, the decrease over this period for no-stated-theory interventions did not reach statistical significance. It is worth noting, however, that the effect sizes for these two pools of studies were quite similar—indeed, there was no significant difference between theory-based and no-stated-theory interventions with regard to changes in PA over this timeframe. In addition, the effect sizes that ultimately emerged from baseline to follow-up were very similar for theory-based (d = 0.30) and no-stated-theory (d = 0.28) interventions (again, with no significant difference between these two pools of studies over this timeframe). Taken together, these results seem to indicate that grounding an intervention within an explicitly stated theory did not produce greater changes in PA—in terms of both its initial uptake as well as longer-term maintenance—compared to interventions that were not explicitly guided by a theory. This finding aligns with a growing body of research ([see 6,10,11]), that suggests theory-based interventions are no more effective than interventions without a stated theory in changing PA behavior.[2] This might be due to similarities in the use of many BCTs between these two types of interventions. Specifically, it has been shown that no-stated-theory PA interventions are equally likely as theory-based interventions to use several BCT clusters (e.g., shaping knowledge, feedback and monitoring, repetition and substitution).[6] Hence, although theories might serve as a useful guide in the refinement and delivery of an intervention's content and processes, it could be argued that the main "drivers" of PA behavior change are the BCTs themselves.

With regard to the inclusion of BCTs within an intervention, there appeared to be a consistent pattern of significant improvements in PA from baseline to post-intervention—with effect sizes typically in the small-to-medium range—followed by significant decreases from post-intervention to follow-up—with effect sizes typically in the small range. There were, however, two exceptions to this finding. Specifically, among interventions that included comparison of outcomes or self-belief, there were significant small increases in PA from baseline to post-intervention but no changes from post-intervention to follow-up. Follow-up analyses confirmed that the effect sizes from post-intervention to follow-up were significantly different among interventions that included these two BCT clusters relative to those that did not. These results suggest that although these two BCT clusters might not appear as effective as others in the short-term, ultimately their effects from baseline to the longer-term follow-up period are comparable to those other BCT clusters. Moreover, since there were no significant decreases in PA following the conclusion of an intervention, it is possible that including comparison of outcomes and self-belief alongside the other BCT clusters that showed relatively greater improvements from baseline to post-intervention (e.g., comparison of behavior, rewards and threat) might be a useful strategy for maintaining those PA improvements. Unfortunately, we were precluded from examining the potential effects of combining BCT clusters due to limitations in statistical power. Thus, future research remains warranted to determine whether certain combinations of BCT clusters (especially those that include comparison of outcomes and self-belief) might maximize the uptake and maintenance of PA.

It should be noted that it is unclear from the current review why/how these two BCT clusters might assist in the maintenance of PA. However, recent research examining the links between BCTs and mechanisms of action (i.e., the processes through which behavior changes) may offer some insight.[20] Specifically, the processes that might explain how comparison of outcomes influences PA behavior include improvements in one's attitudes towards PA, motivation to be active, and beliefs about the consequences of PA behavior (or non-behavior).[20] For self-belief, these processes include improvements in one's motivation and perceived behavioral capabilities to be active.[20] Future work appears warranted to examine whether these two BCT clusters indeed have a "protective" effect in maintaining PA improvements, and whether those protective effects could be explained by their associated mechanisms of action. In addition, future research could consider the combined effects of these mechanisms of action and the three groups of moderators examined in our meta-analysis (time, theory, and BCT clusters). For example, the relationship between a BCT cluster and mechanism of action may vary based on timeframe (e.g., from pre- to post-intervention compared to post-intervention to follow-up) and/or inclusion of a theory within the PA intervention. Such work could help researchers better identify how exactly these moderators interact and exert their influence on PA behavior.

The main clinical implication from our findings is that the observed improvements initially found in PA interventions are likely to drop in the absence of the intervention. Our findings underscore the difficulty in sustaining PA behavior change. From a practical clinical perspective, those concerned with intervention need to consider the supports that can be provided both during the PA intervention and when the PA intervention concludes to assist with the maintenance of health-enhancing PA behaviors. It should also be noted that in many studies, it is difficult to determine whether interventionists merely include BCTs as part of the intervention or that participants actually learned how to incorporate these behavior change strategies themselves. Although admittedly speculative, perhaps having participants actively learn how to use BCTs once the intervention concludes could help promote PA maintenance to a greater extent than when fully reliant on another person (i.e., interventionist) to deliver those BCTs. For example, rather than simply having participants create PA goals over the course of an intervention, the interventionist could also teach participants how to effectively incorporate goal setting once the intervention has ended. Moreover, our findings provide evidence that incorporating BCTs that target comparison of outcomes and/or self-belief into an intervention hold particular promise in helping sustain the improvements that participants have made over the course of the intervention. Comparison of outcomes include encouraging individuals to compare the reasons between wanting versus not wanting to change their PA behavior, as well as imagining future outcomes associated with changing versus not changing their PA.[3] Targeting self-belief involves building individuals' confidence in their abilities to perform PA using strategies such as self-talk, verbal persuasion, imaginative practice, and focusing on past successes related to PA.[3] Future research could also examine the usefulness of including additional support following the conclusion of an intervention. At present, interventions within behavioral medicine studies typically conclude rather abruptly (i.e., at the end of the scheduled duration of the study). Thus, it could be valuable to examine how practitioners can best assist participants to incorporate the necessary self-regulatory skills needed to sustain regular physical activity once the various supports offered by a formal intervention are no longer available/accessible.

Interestingly, it should be noted that comparison of outcomes and self-belief seem to be included much less frequently in PA interventions compared to many other BCT clusters. Specifically, among the 38 interventions included in the subgroup analyses, comparison of outcomes appeared in 12 (32%) of the interventions while self-belief appeared in 11 (29%). By comparison, goals and planning—the most frequently utilized BCT cluster—appeared in 33 interventions (87%). Hence, there appear to be some missed opportunities to utilize BCTs that might help sustain the PA improvements that are initially derived from an intervention. Relatedly, the effects of several BCTs—including association (k = 4), scheduled consequences (*k* = 4), antecedents (*k* = 4), identity (*k* = 3), regulation (*k* = 1), and covert learning (*k* = 1)—were based on an even smaller number of interventions (and, as such, should be interpreted with caution).[16] These disparities imply that the effectiveness of certain BCT clusters (e.g., goals and planning, shaping knowledge) are well-known as they have been tested quite extensively; in contrast, the effects of several other BCT clusters are much less clear, as they have received comparatively little attention. Nevertheless, given that some of these lesser-studied BCT clusters—including identity, association, and antecedents—appeared to result in the maintenance of PA from post-intervention to follow-up (after deriving pre- to post-intervention improvements in physical activity), it would seem that these clusters may be especially worthy of future investigation. This recommendation may be particularly relevant for association, as this was the only BCT cluster that showed significant differences in effect sizes from pre-intervention to follow-up between interventions that included the cluster compared to those that did not.

CONCLUSION

Based on the results from this review and others,[6,8,9] it is evident that behavior change interventions can produce significant improvements in PA over the course of the intervention. By comparison, much less research has been devoted to examining the extent to which those improvements in PA are sustained following the conclusion of an intervention. Specifically, we found that fewer than one in five studies from the previous review upon which this current meta-analysis was based included a measure of PA at a follow-up timepoint.[6] Moreover, the bulk of the evidence from studies that examined PA over the longer term suggests that the initial improvements derived from interventions tend to dissipate over time once the interventions ended. Hence, as a field, we seem to have a decent understanding of how to enhance the *uptake* of PA. To advance this area of research, there now needs to be a much greater emphasis towards understanding how to effectively *sustain* PA improvements over time. Equally, clinical practitioners and interventionists should reflect, and act on, the supports that they provide to individuals to assist with the maintenance of the positive changes they made with their PA post-intervention. Recent (and future) efforts to develop and test theoretical frameworks that specifically focus on the maintenance of health behavior could help optimize interventions that are concerned with supporting long-term physical activity adherence behaviors.[7,21,22] The results of the current review suggest that incorporating two BCT clusters in particular into PA interventions appear to hold promise in helping to maintain PA following the conclusion of the intervention: comparison of outcomes and self-belief. Future work could benefit from a greater consideration of BCT clusters that have received relatively limited attention in PA behavior change research, particularly those that appear to show early promise in helping to sustain PA levels, including association (e.g., classical conditioning techniques that reinforce PA behavior, creating environmental prompts for PA), identity (e.g., serving as a role model to oneself, constructing a new identity that includes being a physically active person), and antecedents (e.g., restructuring one's social environment, increasing exposure to cues that predict PA behavior).

REFERENCES

1. Hallal P, Andersen L, Bull F et al. Global physical activity levels: Surveillance progress, pitfalls, and prospects. *Lancet*. 2012;380(9838):247-257.
2. Rhodes R, Janssen I, Bredin S et al. Physical activity: Health impact, prevalence, correlates and interventions. *Psychol Health*. 2017;32(8):942-975.
3. Michie S, Richardson M, Johnston M et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Ann Behav Med*. 2013;46(1):81-95.
4. Glanz K, Rimer BK. *Theory at a glance: a guide for health promotion practice* (2nd Ed.). Bethesda, MD: National Institute of Health, U.S. Government Printing Office; 2005.
5. Hankonen N, Sutton S, Prevost A et al. Which behavior change techniques are associated with changes in physical activity, diet and body mass index in people with recently diagnosed diabetes?. *Ann Behav Med*. 2014;49(1):7-17.
6. McEwan D, Beauchamp M, Kouvousis C et al. Examining the active ingredients of physical activity interventions underpinned by theory versus no stated theory: A meta-analysis. *Health Psychol Rev*. 2018;13(1):1-17.
7. Kwasnicka D, Dombrowski SU, White M, Sniehotta FF. Theoretical explanations for maintenance of behaviour change: A systematic review of behaviour theories. *Health Psychol Rev*. 2013;10(3):277-296.
8. Murray JM, Brennan SF, French DP et al. Effectiveness of physical activity interventions in achieving behaviour change maintenance in young and middle aged adults: A systematic review and meta-analysis. *Soc Sci Med*. 2017;192:125-133.
9. Amireault S, Godin G, Vezina-Im LA. Determinants of physical activity maintenance: a systematic review and meta-analyses. *Health Psychol Rev*. 2013;7(1):55-91.
10. Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among health adults: Meta-analysis of outcomes. *Am J Pub Health*. 2011;101(4):751-758.
11. Prestwich A, Sniehotta FF, Whittington C et al. Does theory influence the effectiveness of health behaviour interventions? Meta-analysis. *Health Psychol*. 2014;33(5):465-474.
12. McEwan D, Beauchamp M, Rhodes R. Using behaviour change techniques to improve physical activity: a systematic review of the current evidence from controlled interventions. *PROSPERO.* 2018; CRD42018085811. Available at www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018085811
13. Borenstein M, Hedges L, Higgins J, Rothstein H. *Comprehensive meta-analysis, version 2.* Englewood, NJ: Biostat; 2005.
14. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155-159.
15. Higgins JP, Green, S. *Cochrane handbook for systematic reviews of interventions (Vol. 5.1.0)*. Chichester: Wiley-Blackwell; 2008.
16. Borenstein M, Hedges L, Higgins J, Rothstein, H. *Introduction to meta-analysis.* Chichester, England: Wiley; 2009.
17. Rosenberg MS. The file-drawer problem revisited: A general weighted method for calculating fail-safe numbers in meta-analysis. *Evolution*. 2005;59(2):464-468.
18. Lawton BA, Rose SB, Elley CR et al. Exercise on prescription for women aged 40-74 recruited through primary care: two year randomised controlled trial. *BMJ*. 2008;337:a2509.
19. Vandelanotte C, Spathonis KM, Eakin EG, Owen N. Website-delivered physical activity interventions: A review of the literature. *Am J Prev Med.* 2007;33(1):54-64.
20. Carey RN, Connell LE, Jonston M et al. Behavior change techniques and their mechanisms of action: A synthesis of links described in published intervention literature. *Ann Behav Med.* 2018;53(8):693-707.
21. Nigg CR, Borrelli B, Maddock J, Dishman RK. A theory of physical activity maintenance. *Applied Psychol.* 2008;57(4):544-560.
22. Rhodes RE. The evolving understanding of physical activity behavior: A multi-process action control approach. In AJ Elliot (Ed.), *Advances in Motivation Science* (Vol. 4, pp. 171-205). Cambridge, MA: Elsevier Academic Press; 2017.

Table 1.

Subgroup analyses examining the influence of theory and BCT clusters on effect sizes for each timeframe.

	Pre to Post	**Pre to Follow-Up**	**Post to Follow-Up**
Theory			
Theory-based (k = 28)	d = 0.44, p < .001	d = 0.30, p < .001	d = -0.15, p = .012
No stated theory (k = 10)	d = 0.29, p = .018	d = 0.28, p = .003	d = -0.09, p = .120
BCT Cluster			
Goals & Planning (k = 33)	d = 0.44, p < .001, Q = 4.42*	d = 0.28, p < .001, Q = 0.19	d = -0.16, p = .003, Q = 1.86
Shaping Knowledge (k = 29)	d = 0.40, p < .001, Q = 0.09	d = 0.28, p < .001, Q = 0.77	d = -0.14, p = .017, Q = 0.00
Feedback & Monitoring (k = 27)	d = 0.47, p < .001, Q = 2.67	d = 0.31, p < .001, Q = 0.23	d = -0.19, p = .001, Q = 3.25
Natural Consequences (k = 25)	d = 0.49, p < .001, Q = 6.63*	d = 0.29, p < .001, Q = 0.15	d = -0.20, p = .001, Q = 3.85
Social Support (k = 21)	d = 0.45, p < .001, Q = 1.50	d = 0.29, p < .001, Q = 0.09	d = -0.16, p = .014, Q = 0.44
Repetition & Substitution (k = 15)	d = 0.49, p < .001, Q = 0.51	d = 0.38, p < .001, Q = 0.43	d = -0.18, p = .042, Q = 0.40
Rewards & Threat (k = 12)	d = 0.54, p < .001, Q = 1.84	d = 0.34, p < .001, Q = 0.07	d = -0.18, p = .033, Q = 0.52
Comparison of Outcomes (k = 12)	d = 0.32, p = .002, Q = 0.72	d = 0.34, p < .001, Q = 1.06	d = 0.00, p = .977, Q = 3.88*
Comparison of Behavior (k = 11)	d = 0.62, p < .001, Q = 6.43*	d = 0.36, p < .001, Q = 1.10	d = -0.28, p = .003, Q = 3.34
Self-Belief (k = 11)	d = 0.34, p = .001, Q = 0.34	d = 0.30, p < .001, Q = 0.14	d = -0.01, p = .929, Q = 3.95*
Association (k = 4)	d = 0.30[a], p = .085, Q = 0.29	d = 0.27[a], p = .030, Q = 0.00	d = -0.06[a], p = .703, Q = 0.33
Scheduled Consequences (k = 4)	d = 0.85[a], p < .001, Q = 6.33*	d = 0.52[a], p = .001, Q = 2.61	d = -0.44[a], p = .011, Q = 3.57
Antecedents (k = 4)	d = 0.57[a], p = .002, Q = 1.14	d = 0.54[a], p < .001, Q = 3.91*	d = -0.11[a], p = .482, Q = 0.02
Identity (k = 3)	d = 0.48[a], p = .016, Q = 0.23	d = 0.48[a], p = .001, Q = 1.97	d = -0.05[a], p = .793, Q = 0.29
Regulation (k = 1)	--	--	--
Covert Learning (k = 1)	--	--	--

Note. [a] = these effect sizes should be interpreted with caution as they are based on the results from fewer than six interventions.[8] For each BCT cluster, Q-values estimate the differences between the pool of interventions that incorporated that BCT cluster in the intervention versus those that did not; significant differences between these pools of studies are noted by *p < .05, **p < .01, or ***p < .001.