Published in final edited form as:
J Prim Prev. 2014 August ; 35(4): 203–215. doi:10.1007/s10935-014-0352-5.

Impact of physical activity interventions on anthropometric outcomes: Systematic review and meta-analysis

Vicki S. Conn, PhD RN FAAN,
School of Nursing, University of Missouri

Adam Hafdahl, PhD,
ARCH Statistical Consulting, LLC

Lorraine J. Phillips, PhD, RN,
School of Nursing, University of Missouri

Todd M. Ruppar, PhD, RN, GCNS-BC, and
School of Nursing, University of Missouri

Jo-Ana D. Chase, MN, APRN-BC
School of Nursing, University of Missouri

Summary

Considerable research has tested physical activity interventions to prevent and treat overweight and obesity. This comprehensive meta-analysis synthesized the anthropometric effects of supervised exercise interventions and motivational interventions to increase physical activity. Eligible intervention studies included healthy participants with reported anthropometric outcomes (e.g., body mass index). Extensive searching located 54,642 potentially eligible studies. We included data from 535 supervised exercise and 283 motivational interventions in our syntheses, which used random-effects analyses. Exploratory moderator analyses used meta-analytic analogues of ANOVA and regression. We synthesized data from 20,494 participants in supervised exercise and 94,711 undergoing motivational interventions. The overall mean effect sizes (ES, *d*) for treatment vs. control groups in supervised exercise interventions were 0.20 (treatment vs. control within-group comparison) and 0.22 (between-group comparison). The ES of 0.22 represents a post-intervention body mass index of 26.7 kg/m^2 for treatment participants relative to 27.7 kg/m^2 for controls. The corresponding mean ES for motivational interventions was significantly smaller ($d = .09$ for between group, $d = 0.10$ for treatment vs. control within-group). Control group within-group comparisons revealed slightly worsening anthropometric outcomes during study participation ($d = -0.03$ to $d = -.04$). Moderator analyses identified potential variables for future research. These findings document significant improvements in anthropometric effects from both supervised exercise and motivational interventions.

Corresponding author: Vicki S. Conn PhD RN FAAN, S317 School of Nursing, University of Missouri, Columbia, MO 65211, 573 882 0231 (office), 573 884 4544 (fax), 573 446 1511 (cell), conn@missouri.edu.

Potential conflicts of interest:
none

Keywords

Exercise; Weight; Overweight; Quantitative Synthesis

Introduction

Researchers have consistently documented alarming rates of excess weight and obesity (Wang & Beydoun, 2007) in the U.S. population. Excess weight has been linked to multiple illnesses, including type 2 diabetes, cardiovascular diseases, asthma, osteoarthritis, and several cancers (Guh et al., 2009; Suzuki, Orsini, Saji, Key, & Wolk, 2009). Preventing weight gain in normal weight adults, or facilitating small weight reductions in overweight adults, may produce health benefits (Donnelly et al., 2009; Fox & Hillsdon, 2007; Jakicic, 2009; Wareham, van Sluijs, & Ekelund, 2005). Inadequate physical activity (PA) is one cause of the excess weight epidemic, and proper PA is a potential treatment for reducing weight and maintaining weight loss (Jakicic & Otto, 2005; Wareham, van Sluijs, & Ekelund, 2005). PA is an attractive intervention because it may have health benefits beyond weight control (Jakicic, 2009). Numerous PA intervention trials have examined anthropometric outcomes such as weight, body mass index (BMI), and percent body fat. The proliferation of such studies suggests the need to quantitatively synthesize this body of work to improve outcomes by identifying characteristics of effective interventions.

Some evidence suggests that interventions combining PA and diet are more effective than diet alone in producing and maintaining weight reduction (Anderson, Konz, Frederich, & Wood, 2001; Avenell et al., 2004; Curioni & Lourenco, 2005; Franz et al., 2007). Evidence is mixed regarding whether PA interventions alone can significantly alter anthropometric outcomes (Franz et al., 2007; Garrow & Summerbell, 1995; Miller, Koceja, & Hamilton, 1997; Ohkawara, Tanaka, Miyachi, Ishikawa-Takata, & Tabata, 2007; Richardson et al., 2008). The most comprehensive meta-analysis of PA's effects on anthropometric measures, published in 1991, included only 53 studies and split findings into several groups; anthropometric effects were linked with gender and baseline anthropometric values (Ballor & Keesey, 1991). Most previous meta-analyses of anthropometric outcomes of PA interventions synthesized few studies and did not conduct moderator analyses (Franz et al., 2007; Miller et al., 1997; Norris et al., 2005; Ohkawara et al., 2007; Shaw, Gennat, O'Rourke, & Del Mar, 2006). Other meta-analyses did not separate interventions for diet and PA (Galani & Schneider, 2007), and some focused on clinical populations (Avenell et al., 2004; Norris et al., 2005). Important questions remain regarding the magnitude of PA intervention effects on anthropometric outcomes. Previous syntheses addressed neither supervised exercise interventions that manipulated exercise dose (henceforth "supervised exercise"), nor interventions designed to motivate participants to increase PA (henceforth "motivational interventions"). It is important to distinguish programs that directly manipulate PA dose from those that simply recommend increased PA. In addition, moderator analyses are needed that examine sample and intervention characteristics associated with better anthropometric effects.

This study extends previous syntheses in three ways. First, we comprehensively searched for PA studies with anthropometric outcomes. Second, we examined both supervised exercise and motivational interventions. Third, we conducted moderator analyses to determine which characteristics of sample, design, or intervention were associated with effects. Our primary research questions were: (1) What is the overall effect of supervised exercise interventions on anthropometric outcomes? (2) How do the anthropometric effects of supervised exercise vary depending on sample, design, or intervention characteristics? (3) What is the overall effect of motivational interventions on anthropometric outcomes? (4) How do the anthropometric effects of motivational interventions vary depending on sample, design, or intervention attributes?

Methods

We used widely-accepted research synthesis and meta-analysis methods, including PRIMSA guidelines, to locate and retrieve potential studies, assess eligibility, reliably code data, meta-analyze primary study results, and interpret findings (Cooper, Hedges, & Valentine, 2009; Liberati et al., 2009). The protocol we used is available from the corresponding author.

Search Strategies

We utilized multiple search strategies to limit bias from narrow searches (White, 2009). An experienced reference librarian performed searches in 13 databases (e.g., MEDLINE, SportDiscus). We conducted searches from 1960 forward because computerized databases rarely contain papers published before that date. We searched 36 research registers (e.g., CRISP, National Health Service Clinical Trials Register) to identify potential studies (Reed & Baxter, 2009). We used comprehensive search terms (PA: exercise, exercise therapy, exertion, physical activity, physical fitness, physical education and training, walking; intervention: adherence, behavior therapy, clinical trial, compliance, counseling, evaluation, evaluation study, evidence-based medicine, health care evaluation, health behavior, health education, health promotion, intervention, outcome and process assessment, patient education, program, program development, program evaluation, self-care, treatment outcome, validation study). Staff hand-searched 82 journals that published more than two eligible studies,, because computerized database searches generally miss some eligible studies (Reed & Baxter, 2009). We also conducted computerized database searches on senior authors and principal investigators of eligible studies, and performed ancestry searches on reference lists and review articles.

Inclusion Criteria

We included studies if they reported anthropometric outcomes following either supervised exercise with verified exercise dose or motivational interventions. Supervised exercise studies used research staff-controlled exercise sessions at centralized exercise locations to test the effects of a verified exercise dose. To ensure that the outcomes were based on a verified exercise dose, we included supervised exercise studies only if anthropometric outcomes were measured within seven days of completion of the most recent regulated exercise session. Primary study investigators designed motivational and educational interventions to convince subjects to change their exercise behavior. Motivational interventions did not include a verified exercise dose within the previous six months.

We synthesized primary studies if they included healthy adult samples, regardless of sample weight characteristics. We included studies that described samples as healthy or without acute or chronic illnesses. Eligible anthropometric outcomes included weight, BMI, percent body fat, and central obesity (i.e., waist-to-hip ratio). We included studies with diverse interventions to represent existing literature and facilitate moderator analyses of the effects of intervention variations. To adequately represent extant research, we included primary studies that exclusively targeted PA behavior and studies that focused on changing both PA behavior and diet. We used moderator analyses to address differences between studies that combined PA and diet and studies that only addressed PA behavior. We included both published and unpublished studies because the statistical significance of findings is the most consistent difference between published and unpublished reports and because publication status is an inadequate proxy measure of study quality (Rothstein & Hopewell, 2009). We included small sample studies because, though they may lack statistical power, they do contribute to effect size (ES) estimates. We weighted ES estimates so more precise estimates (e.g., due to larger sample sizes) would be proportionally given more influence on our findings.

Eligibility Determinations

Full-time project research specialists were employed to screen searches conducted by reference librarians, conduct other searches, and determine eligibility. Employed research specialists were necessary because of the size of the parent project; 54,642 studies were considered for inclusion. All research specialists had graduate degrees in health or behavioral sciences. We applied a staged eligibility protocol. Research specialists determined initial eligibility by the presence of an exercise intervention, which constituted either a supervised exercise or a motivational intervention. Research specialists initiated this stage when reviewing potential study abstracts and confirmed with full reports. Research specialists always performed the next stage of eligibility on full reports to determine the eligibility of primary study samples as healthy adults. Research specialist and the principal investigator (VSC) determined the third stage of eligibility by the presence of outcome measures for the parent project and the adequacy of findings for calculating ESs. The principal investigator or other co-investigators evaluated potential primary studies with unclear eligibility.

Primary Study Quality and Risk of Bias

The scientific rigor of the studies we reviewed varied dramatically. Strategies meta-analysts use to address quality include *a priori* limitations to including studies with quality problems and *post hoc* approaches that consider quality as an empirical question (Cooper, 2009; Valentine, 2009). Most syntheses apply a blend of approaches (Valentine, 2009), and ours used both. Design is an important component of study quality. We partially addressed design bias by reporting ESs separately for two-group post-intervention comparisons, two-group pre-post comparisons, and treatment single-group comparisons. We reported control single-group comparisons to explore potential bias from research study participation. We excluded studies without control groups from the main analyses, and presented findings from single-group studies as ancillary findings.

We considered the impact of subject randomization as an empirical question. We conducted a sensitivity analysis on the method of assignment of subjects to groups. Attrition is an important quality feature in this area of research. We used moderator analyses to assess the potential link between attrition and study outcomes. We statistically managed another common scientific rigor attribute, sample size. We included small sample studies because, though they may lack statistical power, they do contribute to ES estimates. We weighted ES estimates so that more precise estimates (e.g., due to larger sample sizes) would be given proportionally more influence on our findings. Although publication status has major limitations as a proxy measure of study quality, we conducted moderator analyses on publication status as a form of sensitivity analysis.

To avoid the bias introduced by including studies with larger ESs, which are often easier to locate, we used intensive and extensive search strategies. We managed potential reporting bias within studies by coding anthropometric measures determined *a priori*. We included both published and unpublished studies because the statistical significance of findings is the most consistent difference between published and unpublished reports and because publication status is an inadequate proxy measure of study quality (Rothstein & Hopewell, 2009). We assessed publication bias using multiple methods described in Table 1.

Data Extraction, Management, and Analyses

To assess report, experimental, sample, intervention, and outcome data, we developed a coding frame from previous related meta-analyses, review papers, and extensive examinations of primary reports (Lipsey, 2009). We pilot tested the coding frame on 30 studies and revised it prior to full implementation. We coded data at the micro-level to enhance reliability and validity (Orwin & Vevea, 2009).

Two extensively trained coders independently extracted data from each eligible report. A doctorally prepared researcher independently verified all ES data. To ensure valid data, we compared independently assessed codes for each variable to achieve full consensus.

We analyzed separately studies of anthropometric outcomes of supervised exercise and motivational interventions. We calculated a standardized mean difference ($\hat{\mu}_\delta$) ES for each comparison (Morris & DeShon, 2002). Details of the analyses are provided in Table 1.

Results

Comprehensive searching yielded 54,642 reports that we considered for inclusion. Figure 1 depicts the flow of potential studies through the study. We coded eligible data from 535 supervised exercise and 283 motivational intervention research reports. Among the supervised exercise studies with outcomes measured within seven days of completing the supervised exercise, 81studies contributed an ES to at least one analysis and comprised an intervention that included some motivational component. Sixteen studies had an intervention classified as motivational that also had a supervised exercise component completed at least

180 days prior to outcome measurement, and contributed an ES to at least one analysis. Dissertations and presentations contributed unpublished studies (*s*=66; *s* denotes the number of reports, *k* indicates the number of comparisons). Primary study authors published 236 reports in 2000 or later. The list of primary studies is available in electronic supplementary material.

Characteristics of Primary Studies

We calculated supervised exercise ESs from data pertaining to 20,494 participants. Between-group analyses included 302 comparisons of 8,927 participants (*s*=208). Most (*k*=219) were randomized controlled trails. Treatment within-group quasi-experimental comparisons used data from 16,806 participants (*k*=763, *s*=530). We calculated motivational intervention ESs from data for 94,711 participants. Between-group comparison analyses included 76 comparisons of 45,992 participants (*s*=63). Forty-three of these studies were randomized controlled trials. Treatment within-group quasi-experimental comparisons used data from 56,258 participants (*k*=436, *s*=275).

Table 2 lists key characteristics of primary studies. Primary study authors reported median post-test sample size of 22 participants for supervised exercise studies and 56 participants for motivational interventions. Authors reported median of mean ages of 37 years for supervised exercise and 43 years for motivational interventions. The primary study samples included a median of 55% women for supervised exercise and 81% for motivational studies. Only 48 supervised exercise and 79 motivational intervention studies reported ethnicity, with a median percentage ethnic minority of 15% or less among the studies providing this information. Baseline BMI values (mdn=27 to 30 kg/m^2) documented the prevalence of overweight subjects in samples, with a BMI over 25 kg/m^2. We documented attrition among supervised exercise studies (mdn=5% to 8%) and motivational intervention studies (mdn=15% to 18%).

Overall Effects of Supervised Exercise Interventions

We excluded identified outlier ESs from the main analyses. For supervised exercise, we omitted the following: between-group (5, or 2%, when ignoring multiple-treatment dependence; 6, or 2%, otherwise), two-group within-group (9, or 3%), treatment within-group (44, or 5%), and control within-group ESs (5, or 2%).

Table 3 shows the estimated overall mean ESs, confidence intervals, and associated statistics calculated for anthropometric outcomes with outliers excluded (see Table 4 in the supplementary materials reports results with outliers included). Supervised exercise significantly improved anthropometric effects for between-group comparisons ($\hat{\mu}_\delta = 0.22$) for analyses with and without outliers. In terms of BMI, which had an average standard deviation of 4.6, the between-group mean ES of 0.22 for supervised exercise is consistent with a mean difference of 1.0 between treatment and control participants. For instance, with a mean BMI difference of 1.0 treatment participants would end studies with a mean BMI of 26.7 kg/m^2, whereas control participants' mean BMI would be 27.7 kg/m^2. In terms of weight, the between-group comparison mean ES of 0.22 is consistent with a mean difference of 2.4 kg between treatment and control participants (average *SD*=10.9).

Supervised exercise significantly improved anthropometric effects for treatment vs. control within-group comparisons (outliers excluded: $\hat{\mu}_\delta = 0.20$; outliers included: $\hat{\mu}_\delta = 0.21$). The treatment within-group comparisons also documented significant improved anthropometric effects (outliers excluded: $\hat{\mu}_\delta = 0.20$, outliers included: $\hat{\mu}_\delta = 0.22$). In contrast, control within-group comparisons revealed statistically significant worse anthropometric outcomes at final data collection compared to baseline assessment (outliers excluded: $\hat{\mu}_\delta = -0.23$, outliers included: $\hat{\mu}_\delta = -0.02$). Findings from heterogeneity analyses (Q and I^2) suggest substantial variation in true ESs among studies for the between-group and treatment within-group comparisons, but homogeneity for the two other types of comparisons.

Supervised Exercise Moderator Analyses

We conducted moderator analyses on between-group data with outliers excluded for supervised exercise (see Tables 5 and 6 in electronic supplementary material).

Dichotomous Moderators—The difference between studies that included participants without pre-intervention exercise behavior ($\hat{\mu}_\delta = 0.24$) and those whose participants reported pre-intervention exercise ($\hat{\mu}_\delta = 0.15$) was not statistically significant (see Table 5 in electronic supplementary material). Interventions that targeted both exercise and diet behavior did not report significantly different anthropometric effects ($\hat{\mu}_\delta = 0.29$) than studies targeting only PA behavior ($\hat{\mu}_\delta = 0.22$). Supervised exercise interventions that included both aerobic and resistance exercise did not yield significantly different anthropometric effects ($\hat{\mu}_\delta = 0.17$) than interventions using only endurance exercise ($\hat{\mu}_\delta = 0.24$). Our analyses of other potential dichotomous moderators (publication status, random subject assignment) did not yield significant Q_{between} values.

Continuous Moderators—Exercise dose was an important variable (see Table 6 in electronic supplementary material). Studies with more exercise sessions reported larger improvements in anthropometrics. We observed a similar pattern for the total minutes of supervised exercise in both the cubic and linear analyses. Further analyses indicated that the total number of sessions was the important component of the total minutes of supervised exercise. In contrast, neither the number of minutes of supervised exercise per session, nor the frequency of exercise sessions per week, was a significant moderator of anthropometric effects. Analyses comparing low-, moderate-, and high-intensity exercise did not yield statistically significant differences by intensity. Neither sample age, proportion female, proportion minority, attrition, baseline BMI, baseline weight, nor publication year was a significant moderator in the mixed-effects analysis.

Overall Effects of Motivational Interventions

For motivational interventions we omitted identified outliers from the main analyses: between-group (5, or 6%), treatment vs. control within-group (6, or 8%), treatment within-group (18, or 4%), and control within-group ESs (8, or 12%).

Motivational interventions significantly improved anthropometric outcomes. Table 3 shows the estimated overall mean ESs, confidence intervals, and associated statistics calculated for anthropometric outcomes of motivational interventions with outliers excluded (analyses with outliers included are in Table 4 of the supplementary materials). The mean ES for between-group comparisons was 0.09 for the analyses with outliers excluded and 0.17 with outliers included. In terms of BMI, the between-group mean ES of 0.09 for motivational interventions is consistent with a mean difference of 0.4 between treatment and control participants (average *SD*=4.4). For instance, treatment participants would end studies with a mean BMI of 27.0 kg/m^2 while control participants' mean BMI would be 27.4 kg/m^2. In terms of weight, the between-group mean ES of 0.10 is consistent with a mean difference of 1.3 kg between treatment and control participants.

For treatment vs. control within-group comparisons, the mean ES was 0.10 with outliers excluded and 0.15 with outliers included. The mean treatment within-group ES was 0.20 for the analyses without outliers and 0.21 for the analyses including outliers. The control within-group comparisons documented significantly worse anthropometric outcomes at final assessment relative to baseline values (outliers excluded: $\hat{\mu}_\delta = -0.04$, outliers included: $\hat{\mu}_\delta = -0.05$). All of the ESs for motivational interventions were significantly heterogeneous.

Motivational Intervention Moderator Analyses

We conducted moderator analyses on between-group data with outliers excluded for motivational interventions (see Tables 7 and 8 in electronic supplementary material).

Dichotomous Moderators—Several of the dichotomous moderator analyses should be interpreted cautiously, given the small sample sizes (*k*) for some variables. Unpublished studies ($\hat{\mu}_\delta = 0.32$) reported larger anthropometric effects than published studies ($\hat{\mu}_\delta = 0.08$). Neither sample nor design characteristics were related to anthropometric effects. Motivational interventions included diverse strategies to increase PA behavior. Studies that reported using the transtheoretical model to design interventions reported larger ESs ($\hat{\mu}_\delta = 0.21$) than studies using social cognitive theory ($\hat{\mu}_\delta = 0.07$). There were no differences in anthropometric effects between studies that used behavioral strategies and those using cognitive strategies. Studies that described their intervention as "behavior modification" reported poorer anthropometric effects ($\hat{\mu}_\delta = 0.00$). than studies not using this descriptor ($\hat{\mu}_\delta = 0.11$) Interventions that did not recommend walking as PA ($\hat{\mu}_\delta = 0.13$) improved anthropometric effects more than those that did recommend it ($\hat{\mu}_\delta = 0.04$). The difference between interventions that recommended low-intensity PA ($\hat{\mu}_\delta = 0.10$) and moderate- or high-intensity PA ($\hat{\mu}_\delta = 0.19$) was not statistically significant. Studies reported similar ESs regardless of whether they targeted PA behavior only ($\hat{\mu}_\delta = 0.08$) or targeted both PA behavior and diet ($\hat{\mu}_\delta = 0.10$). The difference between studies with face-to-face intervention delivery and those with mediated delivery was not statistically significant. Linking interventions to worksites or primary care did not significantly change anthropometric effects. Other dichotomous potential moderators were unrelated to anthropometric effects (see Table 7 in electronic supplementary material).

Continuous Moderators—Age was significant in the cubic analysis but not in the linear analysis (see Table 8 in electronic supplementary material). The mean ES was lowest for samples whose mean age was in the mid to late 20s or in the late 40s to mid 50s, and was highest for those in the 30s or early 60s. Neither baseline weight nor BMI was a significant moderator of ESs. Studies that measured outcomes around 35 to 300 days reported the best anthropometric effects. Studies that recommended about 100 to 250 minutes of PA per week reported lower anthropometric ESs than studies that recommended fewer or more minutes.

The mixed-effects meta-regression analyses of baseline mean BMI and weight for supervised PA and motivational interventions provide weak evidence for an association between baseline mean and mean posttest ES. All four analyses indicate a larger positive intervention effect for samples with higher baseline mean, especially in regards to weight, but only one of them is statistically significant (see Table 9 in electronic supplementary material: this table depicts ESs at three quartiles for baseline BMI and weight).

Comparison of Supervised and Motivational Interventions and Evidence of Publication Bias

We compared anthropometric between-group mean ESs of supervised exercise and motivational interventions. Supervised exercise was significantly more effective than motivational interventions in improving anthropometric outcomes (z=3.70, p=.0002).

The results of selection function estimation and funnel-plot asymmetry assessment suggest no evidence of publication bias for between-group ESs from supervised exercise. For the other five combinations of ES type and intervention type, we found some evidence of publication bias. Adjusting for this publication bias using selection-function methods increased the mean ES for intervention ESs (i.e., between-group and treatment within-group for supervised exercise, treatment within-group for motivational interventions).

Discussion

Main Findings

Obesity is projected to become the leading determinant of preventable burden of disease (Lemmens, Oenema, Klepp, Henriksen, & Brug, 2008). This meta-analysis is the first comprehensive synthesis of the effects of supervised exercise and motivational interventions on anthropometric outcomes. The supervised exercise ES finding equivalent to 2.4 kg is similar to that of a smaller meta-analysis of combined diet and exercise interventions in overweight adults that reported a 2.19 kg (k=13) weight change (Galani & Schneider, 2007). Exercise-only interventions have reported similar changes: 2.8 kg (k=7; Garrow & Summerbell, 1995), 1.27 kg (k=9; Richardson et al., 2008), and 2.9 kg (k=90; Miller et al., 1997). Larger weight losses (7.47 kg to 14.99 kg) were reported in a meta-analysis of two-year structured programs with very low-energy diets (<800 kcal/day) in overweight samples (Anderson et al., 2001). The variations in weight loss likely reflect partially non-overlapping primary studies, usually resulting from limited searches. Our comprehensive search strategies likely retrieved more studies with smaller ESs than previously reported meta-analyses with larger ESs, because more obscure studies are likely to have smaller ESs (Borenstein, Hedges, Higgins, & Rothstein, 2009; Rothstein & Hopewell, 2009; Sutton, 2009). This comprehensive meta-analysis thus provides more stable estimates of ESs.

The accumulated improvements in anthropometric outcomes across years of PA could be large. In this meta-analysis interventions extended approximately 14 weeks. Weight loss could be greater over longer follow-up periods (Anderson et al., 2001; Curioni & Lourenco, 2005; Franz et al. 2007). Small weight losses may be important. The early loss of visceral fat may have significant health benefits (Atlantis, Barnes, & Singh, 2006; de Koning, Merchant, Pogue, & Anand, 2007; Lee, Huxley, Wildman, & Woodward, 2008). Attention to visceral fat assessments in future research may allow future meta-analyses to quantify the effects of PA on central obesity. Exercise confers health benefits regardless of weight loss (Shaw et al., 2006). Indeed, fitness is a better predictor of mortality than smoking status, blood pressure, or cholesterol levels (Wei et al., 1999). These important health benefits of PA can be used to encourage adults to engage in regular PA, regardless of changes in anthropometric outcomes.

Several possible explanations exist for the modest anthropometric effects in supervised exercise studies that we have reported. Participants may have reduced their overall PA or increased their food intake, or the exercise stimulus may have been insufficient (Thomas et al., 2012). Sedentary participants may have increased their muscle mass without changing weight. Future studies could address this issue by employing accurate measures of body composition.

Although these studies had limited follow-up, control participants gained a small amount of weight while treatment participants reduced weight. The control sample weight gain is consistent with previous findings regarding the importance of exercise in preventing weight gain (Anderson et al., 2001). Epidemiological data suggest that normal-weight women require 60 minutes of moderate-intensity activity every day to avoid weight gain (Lee, Djousse, Sesso, Wang, & Buring, 2010). Projecting weight changes over several years could result in highly clinically meaningful improvements (Ballor & Keesey,1991; Ohkawara et al., 2007; Shaw et al., 2006). These findings suggest every effort should be made to increase PA among sedentary adults to prevent weight gain. Behavioral strategies, such as self-monitoring PA behavior, cues or prompts to exercise, rewards for PA behavior, specific goal setting, and contracting, which have been found effective in large scale meta-analyses, should be used to increase PA to minimize overweight and obesity (Conn, Hafdahl, Brown, & Brown, 2008; Conn, Hafdahl, & Mehr, 2011; Conn, Valentine, & Cooper, 2002).

Implications of Moderator Findings

Our analyses found no difference between studies that exclusively targeted PA and those that combined PA and diet changes. Previous reviews have reported mixed findings in this regard (Anderson et al., 2001; Avenell et al., 2004; Curioni & Lourenco, 2005; Franz et al., 2007; Miller et al., 1997; Shaw et al., 2006). The possibility that subjects differentially alter other concurrent PA, or may not change their diet, may help explain these findings.. Limited follow-up may obscure possible long term differences between exercise only and exercise plus diet interventions.

The finding that ES was unrelated to baseline weight or BMI was unexpected. Smaller reductions in anthropometric outcomes might be expected for normal-weight samples. A previous meta-analysis of 30 studies reported slightly larger weight loss among obese subjects (3.49 kg) as compared to overweight subjects (2.19; Galani & Schneider, 2007). An early attempt to synthesize studies without standard meta-analytic estimation of ESs suggested greater weight loss among overweight relative to normal weight subjects (Epstein & Wing, 1980). Most previous meta-analyses have not compared normal weight to overweight samples. It is possible that the similar improvements across baseline weight and BMI result from short study duration. Research that follows samples with verified exercise dose over multiple years may find greater improvements in anthropometric outcomes among overweight and obese adults than among normal weight subjects.

This study was designed as a comprehensive review. Findings confirmed the expected heterogeneity. The moderator analyses identified some study characteristics linked with effects. In supervised exercise studies, new exercisers (those without a history of exercise) lost more weight than routine exercisers. The exercise dose, specifically the number of exercise sessions, was also a significant moderator. This finding suggests that healthy samples should be encouraged to engage in long-term frequent exercise sessions to achieve desired outcomes.

Limitations

Study limitations include methods weaknesses specific to individual studies and to meta-analysis. Anthropometric assessments may have measurement error. Participants may experience changes in fat-free mass and visceral adiposity without having weight or BMI changes (Mekala & Tritos, 2009). Some moderator variables were inconsistently reported in a manner that limits confidence in the findings we reported. While individual subject level data would have been useful for some analyses, such as to address whether baseline anthropometric scores affect outcomes, no individual subject level information was used in this project. Important components of study quality and treatment fidelity are poorly reported in primary studies. Moderator analyses are complicated because of potential confounding by other moderators. Separating a target moderator’s effect from that of associated moderators is difficult, especially when substantial amounts of data are missing. Despite comprehensive searches for companion papers to provide adequate data for calculating ESs, we excluded some studies for lack of data. Finally, meta-analysis findings are observational and should generate further research.

Conclusions

This comprehensive meta-analysis documented significant positive anthropometric effects of both supervised exercise interventions and motivational interventions. Although the magnitude of effects may appear small, significant health benefits may accrue when PA is extended over multiple years. Motivational interventions were modestly effective in improving anthropometric outcomes. The pattern of worsening anthropometric outcomes among control participants suggests that one of PA’s most important benefits may be in preventing weight gain. The exploratory moderator analyses suggest directions for future primary research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support provided by a grant from the National Institutes of Health (R01NR009656) to Vicki Conn, principal investigator. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

Anderson JW, Konz EC, Frederich RC, Wood CL. Long-term weight-loss maintenance: A meta-analysis of US studies. American Journal of Clinical Nutrition. 2001; 74(5):579–584. [PubMed: 11684524]

Atlantis E, Barnes EH, Singh MAF. Efficacy of exercise for treating overweight in children and adolescents: A systematic review. International Journal of Obesity. 2006; 30(7):1027–1040. [PubMed: 16534526]

Avenell A, Brown TJ, McGee MA, Campbell MK, Grant AM, Broom J, et al. What interventions should we add to weight reducing diets in adults with obesity? A systematic review of randomized controlled trials of adding drug therapy, exercise, behaviour therapy or combinations of these interventions. Journal of Human Nutrition and Dietetics. 2004; 17(4):293–316. doi:10.1111/j.1365-277X.2004.00530.x. [PubMed: 15250841]

Ballor DL, Keesey RE. A meta-analysis of the factors affecting exercise-induced changes in body mass, fat mass and fat-free mass in males and females. International Journal of Obesity. 1991; 15(11):717–726. [PubMed: 1838100]

Borenstein, M.; Hedges, L.; Higgins, JP.; Rothstein, HR. Introduction to meta-analysis. John Wiley & Sons; United Kingdon: 2009.

Conn VS, Hafdahl AR, Brown SA, Brown LM. Meta-analysis of patient education interventions to increase physical activity among chronically ill adults. Patient Education and Counseling. 2008; 70(2):157–172. [PubMed: 18023128]

Conn VS, Hafdahl AR, Mehr DR. Interventions to increase physical activity among healthy adults: Meta-analysis of outcomes. American Journal of Public Health. 2011; 101(4):751–758. [PubMed: 21330590]

Conn VS, Valentine JC, Cooper HM. Interventions to increase physical activity among aging adults: A meta-analysis. Annals of Behavioral Medicine. 2002; 24(3):190–200. [PubMed: 12173676]

Cooper, H. Research synthesis and meta-analysis: A step-by-step approach. 4th ed.. Sage; Thousand Oaks, CA: 2009.

Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009.

Curioni CC, Lourenco PM. Long-term weight loss after diet and exercise: A systematic review. International Journal of Obesity. 2005; 29(10):1168–1174. [PubMed: 15925949]

de Koning L, Merchant AT, Pogue J, Anand SS. Waist circumference and waist-to-hip ratio as predictors of cardiovascular events: Meta-regression analysis of prospective studies. European Heart Journal. 2007; 28(7):850–856. [PubMed: 17403720]

Donnelly JE, Blair SN, Jakicic JM, Manore MM, Rankin JW, Smith BK, et al. American College of Sports Medicine Position Stand. Appropriate physical activity intervention strategies for weight loss and prevention of weight regain for adults. Medicine and Science in Sports and Exercise. 2009; 41(2):459–471. [PubMed: 19127177]

Epstein LH, Wing RR. Aerobic exercise and weight. Addictive Behaviors. 1980; 5(4):371–388. [PubMed: 7211534]

Fox KR, Hillsdon M. Physical activity and obesity. Obesity Reviews. 2007; 8(Suppl 1):115–121. doi:10.1111/j.1467-789X.2007.00329.x. [PubMed: 17316313]

Franz MJ, VanWormer JJ, Crain AL, Boucher JL, Histon T, Caplan W, et al. Weight-loss outcomes: A systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. Journal of the American Dietetic Association. 2007; 107(10):1755–1767. doi:10.1016/j.jada.2007.07.017. [PubMed: 17904936]

Galani C, Schneider H. Prevention and treatment of obesity with lifestyle interventions: Review and meta-analysis. International Journal of Public Health. 2007; 52(6):348–359. [PubMed: 18368998]

Garrow JS, Summerbell CD. Meta-analysis: Effect of exercise, with or without dieting, on the body composition of overweight subjects. European Journal of Clinical Nutrition. 1995; 49(1):1–10. [PubMed: 7713045]

Gleser LJ, Olkin I. Models for estimating the number of unpublished studies. Statistics in Medicine. 1996; 15(23):2493–2507. [PubMed: 8961459]

Gleser, LJ.; Olkin, I. Stochastically dependent effect sizes. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 357-376.

Guh DP, Zhang W, Bansback N, Amarsi Z, Birmingham CL, Anis AH. The incidence of co-morbidities related to obesity and overweight: A systematic review and meta-analysis. BMC Public Health. 2009; 9:88. [PubMed: 19320986]

Hedges, L.; Olkin, I. Statistical methods for meta-analysis. Academic Press; Orlando, FL: 1985.

Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. British Medical Journal. 2003; 327(7414):557–560. [PubMed: 12958120]

Jakicic J. The effect of physical activity on body weight. Obesity. 2009; 17(Suppl 3):S34–38. [PubMed: 19927144]

Jakicic J, Otto AD. Physical activity considerations for the treatment and prevention of obesity. American Journal of Clinical Nutrition. 2005; 82(1 Suppl):226S–229S. [PubMed: 16002826]

Lee C, Huxley RR, Wildman RP, Woodward M. Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: A meta-analysis. Journal of Clinical Epidemiology. 2008; 61(7):646–653. [PubMed: 18359190]

Lee I, Djousse L, Sesso HD, Wang L, Buring JE. Physical activity and weight gain prevention. Journal of the American Medical Association. 2010; 303(12):1173–1179. [PubMed: 20332403]

Lemmens VE, Oenema A, Klepp KI, Henriksen HB, Brug J. A systematic review of the evidence regarding efficacy of obesity prevention interventions among adults. Obesity Reviews. 2008; 9(5): 446–455. doi:10.1111/j.1467-789X.2008.00468.x. [PubMed: 18298429]

Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. British Medical Journal. 2009; 339:b2700. [PubMed: 19622552]

Lipsey, M. Identifying interesting variables and analysis opportunities. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 147-158.

Mekala KC, Tritos NA. Effects of recombinant human growth hormone therapy in obesity in adults: A meta analysis. Journal of Clinical Endocrinology and Metabolism. 2009; 94(1):130–137. [PubMed: 18940879]

Miller WC, Koceja DM, Hamilton EJ. A meta-analysis of the past 25 years of weight loss research using diet, exercise or diet plus exercise intervention. International Journal of Obesity and Related Metabolic Disorders. 1997; 21(10):941–947. [PubMed: 9347414]

Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. Psychological Methods. 2002; 7(1):105–125. [PubMed: 11928886]

Norris SL, Zhang X, Avenell A, Gregg E, Schmid CH, Lau J. Long-term non-pharmacological weight loss interventions for adults with prediabetes. Cochrane Database of Systematic Reviews. 2005; (2) doi:10.1002/14651858.CD005270.

Ohkawara K, Tanaka S, Miyachi M, Ishikawa-Takata K, Tabata I. A dose-response relation between aerobic exercise and visceral fat reduction: Systematic review of clinical trials. International Journal of Obesity (London). 2007; 31(12):1786–1797. doi:10.1038/sj.ijo.0803683.

Orwin, R.; Vevea, J. Evaluating coding decisions. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 177-203.

Reed, J.; Baxter, P. Using reference databases. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 73-101.

Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. The Annals of Family Medicine. 2008; 6(1):69–77. doi:10.1370/afm.761.

Rosenthal R. The file drawer problem and tolerance for null results. Psychological Bulletin. 1979; 86(3):638–641.

Rothstein, HR.; Hopewell, S. Grey literature. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 103-125.

Shaw K, Gennat H, O'Rourke P, Del Mar C. Exercise for overweight or obesity. Cochrane Database of Systematic Reviews. 2006; (4) doi:10.1002/14651858.CD003817.pub3.

Sterne JA, Egger M. p. b. Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. Journal of Clinical Epidemiology. 2001; 54(10):1046–1055. [PubMed: 11576817]

Sutton, AJ. Publicaton bias. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 435-452.

Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status--a meta-analysis. International Journal of Cancer. 2009; 124(3):698–712.

Thomas DM, Bouchard C, Church T, Slentz C, Kraus WE, Redman LM, et al. Why do individuals not lose more weight from an exercise intervention at a defined dose? An energy balance analysis. Obesity Reviews. 2012; 13(10):835–847. [PubMed: 22681398]

Valentine, J. Judging the quality of primary research. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 129-146.

Vevea JL, Hedges LV. A general linear model for estimating effect size in the presence of publication bias. Psychometrika. 1995; 60(3):419–435.

Wang Y, Beydoun MA. The obesity epidemic in the United States--gender, age, socioeconomic, racial/ethnic, and geographic characteristics: A systematic review and meta-regression analysis. Epidemiologic Reviews. 2007; 29:6–28. [PubMed: 17510091]

Wareham NJ, van Sluijs EMF, Ekelund U. Physical activity and obesity prevention: A review of the current evidence. Proceedings of the Nutrition Society. 2005; 64(2):229–247. [PubMed: 15960868]

Wei M, Kampert JB, Barlow CE, Nichaman MZ, Gibbons LW, Paffenbarger RS Jr. et al. Relationship between low cardiorespiratory fitness and mortality in normal-weight, overweight, and obese men. Journal of the American Medical Association. 1999; 282(16):1547–1553. [PubMed: 10546694]

White, H. Scientific communication and literature retrieval. In: Cooper, H.; Hedges, L.; Valentine, J., editors. The handbook of research synthesis and meta-analysis. 2nd ed.. Russell Sage Foundation; New York: 2009. p. 51-71.

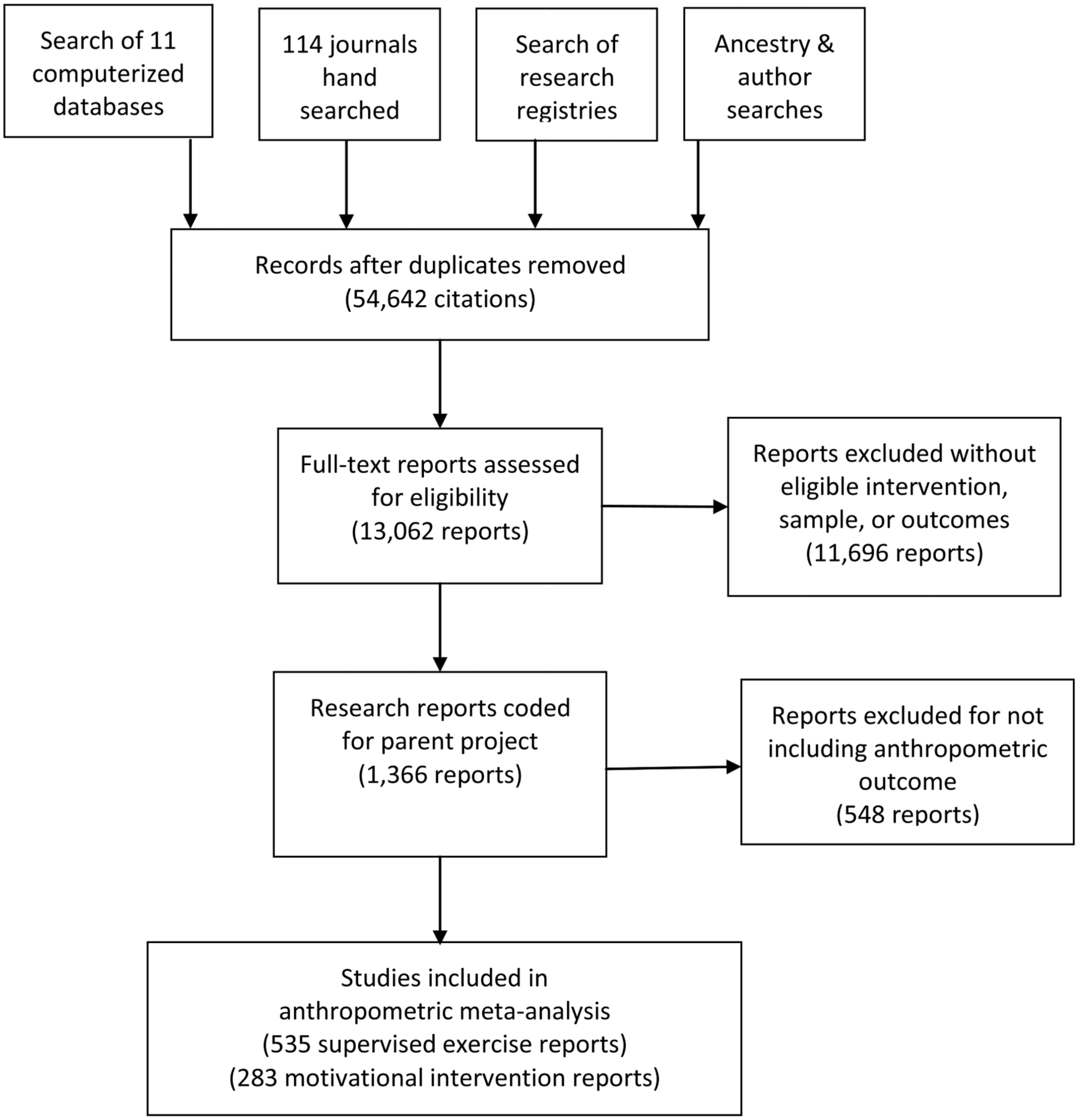

Figure 1.

Flow diagram of potential studies included in meta-analysis.

Table 1

Statistical management of data

Analysis	Rationale or Approach
Standardized mean difference effect size	Standardized unitless measure of difference (each ES weighted by inverse of its sampling variance to give more precise ES more weight; Hedges & Olkin, 1985): • between-group at outcome measure effect • treatment vs. control within-group effect • treatment within-group effect (correlations between pre- and post-intervention scores solicited from primary study authors) • control within-group effect (correlations between pre- and post-intervention scores solicited from primary study authors) Treatment within-group and control within-group ESs complementary evidence
Random-effect model	• ESs based on random-effects models with the between studies variance component σ_δ^2 estimated by weighted method of moments • Random-effects model appropriate when heterogeneity is expected because the model assumes both subject-level sampling error and study-level error
Outlier detection & management	• ES estimates may contain values that do not represent the intended universe of effects • Outlier management included excluding cases that depart substantially from others (Hedges & Olkin, 1985) • Potential outliers identified by omitting each ES and checking for large externally standardized random-effects residuals or substantially reduced measures of heterogeneity • Analyses without outliers emphasized in manuscript • Sensitivity analyses conducted without excluding outliers
Heterogeneity	• Q heterogeneity statistic to test homogeneity • I^2 index of heterogeneity beyond within-study sampling error • Random-effect model for analyses because expected heterogeneity: model assumes both subject-level sampling error and study-level error • Strategies to deal with heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003): ○ Both location parameter and variability parameter reported ○ Reported findings of the random-effects model that assumes heterogeneity ○ Sources of heterogeneity explored by moderator analyses ○ Results interpreted in light of heterogeneity
Multiple treatment groups	Multiple treatment groups compared to same control group included by accounting for dependence resulting from shared control group (Gleser & Olkin, 2009)
ESs as original metric	Mean ES converted to original metric of body mass index and weight: • Meta-analyzed means or SDs from samples that used the measure with the appropriate type of intervention to obtain the hypothetical reference SD and means used to express estimated mean effect sizes in a specific original metric • To determine the mean BMI/weight for the treatment group, the product of the effect size and standard deviation was added to the control group mean
Potential publication bias	Potential publication bias explored using multiple approaches, including estimates of the number of omitted studies, tests of funnel-plot asymmetry, and selection function procedures (Gleser & Olkin, 1996; Rosenthal, 1979; Sterne & Egger, 2001; Sutton, 2009; Vevea & Hedges, 1995)
Exploratory moderator analyses	Moderator analyses conducted on study-level data (not individual subject level data) Continuous moderators:

Analysis	Rationale or Approach
	• Conventional mixed-effects meta-regression procedure to estimate and test unstandardized regression coefficients for both linear and cubic forms of the moderator • Polynomial regression method better detects relationships between ESs and moderators that may be more complex than linear analyses might suggest Dichotomous and categorical moderators: • Meta-analytic analogue of ANOVA • Between-groups heterogeneity statistic (Q_B) to test categorical moderators

Table 2
Characteristics of Primary Studies Included in Anthropometric Meta-Analysis

Characteristic	*s*	Min	Q_1	*Mdn*	Q_3	Max
Supervised exercise						
Mean age (years)	442	18	26	37	53	82
Total post-test sample size per study	535	4	13	22	38	1897
Treatment-group post-test sample size per study	535	4	11	17	29	1897
Comparison-group post-test sample size per study	210	2	8	11	20	102
Percentage attrition from treatment sample	349	0	0	8	23	62
Percentage attrition from comparison sample	150	0	0	5	19	73
Percentage female	515	0	0	55	100	100
Percentage ethnic minority	48	0	0	8	48	100
Mean baseline BMI (kg/m^2) for treatment sample	54	21	25	28	33	46
Mean baseline BMI (kg/m^2) for comparison sample	20	22	23	27	31	36
Minutes of supervised exercise per session	449	7	38	47	60	600
Mean frequency per week of exercise sessions	515	0	3	3	5	36
Total number of supervised exercise sessions	514	2	28	42	72	2028
Motivational interventions						
Mean age (years)	216	18	39	43	48	72
Total posttest sample size per study	283	5	28	56	141	23153
Treatment-group post-test sample size per study	282	5	25	52	129	23153
Comparison-group post-test sample size per study	66	5	19	52	233	5311
Percentage attrition from treatment sample	190	0	6	18	34	79
Percentage attrition from comparison sample	40	0	2	15	30	80
Percentage female	248	0	42	81	100	100
Percentage ethnic minority	79	0	5	15	67	100
Mean baseline BMI (kg/m^2) for treatment sample	84	22	26	30	34	41
Mean baseline BMI (kg/m^2) for comparison sample	28	22	25	27	30	34
Number of intervention strategies	283	0	1	3	5	17
Total number of intervention minutes	61	30	200	720	1440	7415
Recommended minutes of PA per session	89	2	27	30	40	90
Recommended minutes of PA per week	87	11	90	140	175	840

Note. Includes all studies that contributed to primary analyses at least one effect size for any type of comparison. We aggregated samples within studies by summing sample sizes and using weighted mean of other characteristics (weighted by sample size). *s* = number of reports providing data on characteristic; Q_1 = first quartile; Q_3 = third quartile; BMI = body mass index; PA = physical activity.

Table 3

Intervention Effect by Type of Comparison: Random-Effects Estimates and Tests, Outliers Omitted

Comparison type	k	Q	$\hat{\mu}_\delta$	$SE(\hat{\mu}_\delta)$	μ_δ 95% CI	$\hat{\sigma}_\delta$	I^2
			Supervised exercise				
Between-group	301	399 1***	0.22***	0.027	(0.16, 0.27)	0.18	.25
Treatment vs. control within-group	292	234.0	0.20***	0.015	(0.17, ,0.23)	0	0
Treatment within-group	763	1408.5***	0.20***	0.009	(0.18, 0.22)	0.15	.46
Control within-group	218	110.9	−0.03*	0.013	(−0.05, −0.00)	0	0
			Motivational interventions				
Between-group	76	109.3**	0 09***	0.018	(0.06, 0.13)	0.06	.31
Treatment vs. control within-group	67	358.1***	0.10***	0.017	(0.06, 0.13)	0.11	.82
Treatment within-group	436	13927.6***	0.20***	0.012	(0.18, 0.22)	0.23	.97
Control within-group	56	331.0***	−0.04**	0.014	(−0.07, −0.01)	0.08	.83

Note. Under homogeneity (H_0: $\delta_i = \delta$), Q is distributed approximately as chi-square with $df = k - 1$, where k is the number of (possibly dependent) observed effect sizes; this also tests the between-studies variance component, σ_δ^2 (H_0: $\sigma_\delta^2 = 0$). Weighted method of moments used to estimate σ_δ^2.

† $p < .10$.

* $p < .05$.

** $p < .01$.

*** $p <$ 001 (for Q and $\hat{\mu}_\delta$).