Published in final edited form as:

J Evid Based Med. 2017 May ; 10(2): 108–128. doi:10.1111/jebm.12228.

Exercise and BMI z-score in Overweight and Obese Children and Adolescents: A Systematic Review and Network Meta-Analysis of Randomized Trials

Kelley George¹, Kelley Kristi¹, and Pate Russell²

¹Biostatistics, West Virginia University, Morgantown, West Virginia, United States

²Exercise Science, University of South Carolina Ringgold Standard Institution, Columbia, South Carolina, United States

Abstract

Aim—Examine the effects of selected types of exercise (aerobic, strength training, both) on BMI z-score in overweight and obese children and adolescents.

Methods—Randomized exercise intervention trials 4 weeks were included. Studies were retrieved by searching six electronic databases, cross-referencing and expert review. Dual selection and abstraction occurred. Risk of bias and confidence in cumulative evidence were assessed. Network meta-analysis was performed using multivariate random-effects meta-regression models while surface under the cumulative ranking curves were used to calculate a hierarchy of exercise treatments. The number needed to treat (NNT) and percentile improvement (U_3) were also calculated.

Results—Thirty-four studies representing 2,239 participants were included. Median exercise occurred 3 times per week, 50 minutes per session over a 12-week period. Statistically significant reductions in BMI z-score were found for aerobic exercise and combined aerobic and strength exercise, but not strength training alone ($M \pm SD$, 95% CI: aerobic, -0.10, -0.15 to -0.05; aerobic and strength, -0.11, -0.19, -0.03; strength, 0.04, -0.07 to 0.15). Combined aerobic and strength training was ranked best, followed by aerobic exercise and strength training. The NNT was 2 for both aerobic exercise and combined aerobic exercise and strength training. Percentile improvements were 28.8% for aerobic exercise and 31.5% for combined aerobic exercise and strength training. Confidence in effect estimates was ranked as low for aerobic exercise and very low for combined aerobic and strength training as well as strength training.

Conclusions—Aerobic exercise and combined aerobic exercise and strength training are associated with reductions in BMI z-score.

Correspondence: Biostatistics, West Virginia University, Morgantown, West Virginia, United States., Tel: 304-293-6279; Fax: 304-293-5891. gkelley@hsc.wvu.edu.

Conflicts of Interest: None.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as doi: 10.1111/jebm.12228.

Keywords

adolescents; children; exercise; meta-analysis; systematic review; overweight; obesity

Introduction

Overweight and obesity in children and adolescents is a major global health problem. Ng et al. reported that between the years 1980 and 2013, the worldwide prevalence of overweight and obesity in children and adolescents from developed countries has increased from 16.9% to 23.8% in boys and from 16.2% to 22.6% in girls while in developing countries estimated increases of 8.1% to 12.9% (boys) and 8.4% to 13.4% (girls) were reported (1). Between the years 1971 to 1974 and 2011 to 2012, the prevalence of overweight and obesity in the United States has increased from 15.6% to 32.1% among boys and 15.2% to 31.7% in girls (2). Not surprisingly, the estimated costs associated with childhood obesity are high. Finkelstein et al. estimated that the incremental lifetime medical costs of an obese 10-year old child in the United States were \$19,000 greater when compared to a normal weight child of the same age (3). Accounting for the number of obese 10-year-olds in the United States, the lifetime medical costs for this age alone was estimated to be approximately \$14 billion (3).

The deleterious effects of overweight and obesity during childhood and adolescence are both immediate and long-term (4). Specifically, obese children and adolescents are more likely to have risk factors for cardiovascular disease (high cholesterol, high blood pressure, etc.), with approximately 70% possessing at least one risk factor (5). In addition, obese adolescents are more likely to have prediabetes (6). Obese children and adolescents have also been shown to be at a greater risk for bone and joint difficulties, sleep apnea, and social and psychological issues (stigmatization, poor self-esteem, etc.) (7,8). From a long-term perspective, obesity in children and adolescents has been shown to track into adulthood (9-12). As a result, this places them at a greater risk for cardiovascular disease, type 2 diabetes, stroke, several types of cancer, and osteoarthritis during adulthood (4).

While one previous systematic review with meta-analysis reported statistically significant improvements in adiposity as a result of strength training in overweight and obese children and adolescents (13), others that have focused on the effects of exercise as an independent intervention on adiposity as a primary outcome in male and female children and adolescents have reported non-significant changes (14-18). However, all six suffer from one or more of the following potential limitations: (1) inclusion of a small number of studies with exercise as the only intervention (14-16), (2) inclusion of non-randomized trials (13,15), (3) inclusion of children and adolescents who were not overweight or obese (15,17,18), (4) reliance on pairwise versus network meta-analysis that incorporates both direct and indirect evidence (13-18), and (5) absence of an established hierarchy for determining which types of exercise (aerobic, strength training, or both) might be best for improving adiposity (13-18).

The investigative team has previously published meta-analytic research examining the effects of exercise (aerobic, strength training, or both) on body mass index (BMI) z-score (19) and BMI in kg/m^2 (20) in overweight and obese children and adolescents. For both

meta-analyses, statistically significant and practically important reductions of 3-4% were observed (19,20). While these results are encouraging, especially at the population level, they were limited to indirect comparisons that focused on the pooled results of different types of exercise (aerobic, strength training, both) in pairwise meta-analyses (19,20). Network meta-analysis “is a meta-analysis in which multiple treatments (that is, three or more) are compared using both direct comparisons of interventions within randomized controlled trials and indirect comparisons across trials based on a common comparator” (21). In addition, network meta-analysis has the ability provides the opportunity to rank treatments, for example aerobic, strength training, or both, with respect to their effects on the outcome(s) of interest. This is important because practitioners and policymakers want to know which treatments work best and for whom.

To the best of the authors' knowledge, no previous network meta-analysis has examined the effects of aerobic, strength training, or combined aerobic and strength training on BMI *z*-score in overweight and obese children and adolescents. Therefore, the primary objectives of this study were to conduct a systematic review with network meta-analysis of randomized trials to determine the effects of exercise (aerobic, strength training, or both) on BMI *z*-score in overweight and obese children and adolescents as well as provide a ranking of treatments.

Methods

Overview

This study followed the guidelines from the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) extension statement for network meta-analyses of health care interventions (22). The protocol for this systematic review with network meta-analysis has been previously published elsewhere (23) and is registered in PROSPERO (#CRD42015026377).

Study eligibility criteria

The inclusion criteria for this network meta-analysis were as follows: (1) direct evidence from randomized trials that compared two or more exercise interventions (aerobic, strength training, both) or indirect evidence from randomized controlled trials that compared an exercise intervention group to a comparative control group (non-intervention, attention control, usual care, wait-list control, placebo), (2) exercise interventions 4 weeks, (3) male and/or female children and adolescents 2 to 18 years of age, (4) participants overweight or obese, as defined by the original study authors, (5) studies published in any language, (6) published and unpublished studies (dissertations and Master's theses) between January 1, 1990 and September, 2015, and (7) data available or calculable for BMI *z*-score. For the purposes of this study, exercise, aerobic exercise and strength training were defined according to the 2008 Physical Activity Guidelines for Americans (24). Specifically, exercise was defined as movement that is “planned, structured, and repetitive and purposive in the sense that the improvement or maintenance of one or more components of physical fitness is the objective” (24,25), aerobic exercise as “exercise that primarily uses the aerobic energy-producing systems, can improve the capacity and efficiency of these systems, and is effective for improving cardiorespiratory endurance” (24), and strength training as “exercise

training primarily designed to increase skeletal muscle strength, power, endurance, and mass”(24). Studies were limited to randomized trials because it is the only way to control for confounders that are not known or measured as well as the observation that nonrandomized controlled trials tend to overestimate the effects of healthcare interventions (26,27).

Intervention groups that included both exercise and diet performed concurrently were not included unless there was a diet-only group. Four weeks was chosen as the lower cut point for intervention length based on previous research demonstrating improvements in adiposity over this time period in 11-year old girls (28). Participants were limited to overweight and obese children and adolescents as defined by the original study authors because it has been shown that this population is at an increased risk for premature morbidity and mortality throughout their lifetime (29). Studies that took place in any setting were included. Based on our preliminary searching, the first studies that met our inclusion criteria were published in 2004 (30-33). However, to ensure that potentially eligible studies were not missed, the investigators searched back starting with 1990 versus 2004. The rationale for this approach was based on the assumption that no eligible studies would have been published prior to 1990.

Information sources

From the investigators' previous and broad database of studies addressing the effects of exercise on measures of adiposity in overweight and obese children and adolescents published between 1990 and 2012, a search for studies that met the aforementioned eligibility criteria was conducted. In addition, updated searches of the following six databases were conducted for studies in any language that were published and indexed between January 1, 2012 up to September 22, 2015: (1) PubMed, (2) Scopus, (3) Web of Science, (4) Cochrane Central Register of Controlled Trials, (5) CINAHL, (6) Sport Discus. Furthermore, cross-referencing from retrieved studies and reviews was conducted. Finally, the third author, an expert on exercise and obesity in children and adolescents, reviewed the reference list for completeness.

Search strategy

Search strategies were developed using text words as well as medical subject headings (MeSH) associated with the effects of exercise on adiposity in overweight and obese children and adolescents. While the exact terms varied somewhat depending on the database searched, key words included such terms as “exercise”, “random”, “children”, “adolescents”, “overweight” and “obese”. Studies in languages other than English were translated using the freely available Google translate and Babblefish. The second author (KSK) conducted all electronic database searches. A copy of all updated searches can be found in supplementary file 1.

Study selection

All studies were screened, imported into Reference Manager (version 12) (34), and duplicates removed both electronically and manually by the second author (KSK). A copy of the database was then provided to the first author for duplicate screening. The first two authors selected all studies, independent of each other. The full report for each article was obtained for all titles and abstracts that appeared to meet the inclusion criteria or where there

was any uncertainty. Multiple reports of the same study were handled by including the most recently published article as well as drawing from previous reports, assuming similar methods and sample sizes. Neither of the screeners were blinded to the journal titles or to the study authors or institutions. Reasons for exclusion were coded as one or more of the following: (1) inappropriate population, (2) inappropriate intervention, (3) inappropriate comparison(s), (4) inappropriate outcome(s), (5) inappropriate study design, (6) other. Upon completion, both authors met and reviewed their selections. Discrepancies were resolved by consensus. If consensus could not be reached, the third author provided a recommendation. After identifying the final number of studies to be included, the overall precision of the searches was calculated by dividing the number of studies included by the total number of studies screened after removing duplicates (35). The number needed-to-read (NNR) was then calculated as the inverse of the precision (35).

Data abstraction

Prior to the abstraction of data, a codebook that could hold more than 200 items per study was developed using Microsoft Excel 2010 (36). The codebook was developed by the first two authors with input from the third author. The major categories of variables that were coded included (1) study characteristics (author, journal, year, etc.), (2) participant characteristics (age, height, body weight, etc.), and (3) outcome characteristics for BMI z-score and body weight (sample sizes, baseline and post-exercise means and standard deviations, etc.). The first two authors abstracted the data from all studies, independent of each other, using separate codebooks in Microsoft Excel 2010 (36). Upon completion of coding, the codebooks were merged into one primary codebook for review. Both authors then met and reviewed all selections for agreement. Discrepancies were resolved by consensus. If consensus could not be reached, the third author provided a recommendation. Using Cohen's kappa statistic (κ) (37), the overall agreement rate prior to correcting discrepant items was 0.94.

Outcomes and prioritization

The primary outcome in this study was changes in BMI z-score while the secondary outcome was changes in bodyweight in kilograms (kg).

Risk of bias assessment in individual studies

Risk of bias was assessed at the study level using the Cochrane Risk of Bias Assessment instrument (38). The focus was on risk of bias with respect to the primary outcome, changes in BMI z-score. Bias was assessed across six domains: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting and (7) whether or not participants were exercising regularly, as defined by the original study authors, prior to taking part in the study. Each item was classified as having either a high, low, or unclear risk of bias. Since it's probably impossible to blind participants to group assignment in exercise intervention protocols, all studies were classified as high risk of bias for the category "blinding of participants and personnel". Based on previous research, no study was excluded based on the results of the risk of bias assessment (39). All assessments were conducted

following the same procedures as for the abstraction of data. Using Cohen's kappa statistic (κ) (37), the overall agreement rate prior to correcting discrepant items was 0.74.

Data synthesis

Calculation of effect sizes—The primary outcome for this study was changes in BMI *z*-score while the secondary outcome was changes in body weight. For indirect comparisons, changes were calculated by subtracting the change outcome difference in the exercise group minus the change outcome difference in the control group. Variances were calculated from the pooled standard deviations of change scores in the exercise and control groups. If change score standard deviations were not available, these were calculated from either (1) 95% confidence intervals (CI) for change outcome or treatment effect differences, (2) pre and post standard deviation values according to procedures developed by Follmann et al. (40), or (3) data for BMI in $\text{kg}\cdot\text{m}^2$. However, before trying to estimate BMI *z*-score from BMI in kg/m^2 , BMI *z*-score data was requested from investigators. For direct comparisons, i.e., randomized trials with no control group, a post hoc decision was made to use the data augmentation approach described by White since it usually has a negligible effect on results (41). Briefly, studies without a control group arm had one created by using the weighted average of the arm-specific means and standard deviations (41). For studies in which changes in BMI *z*-score and body weight were assessed at multiple intervention time points, for example, 0, 4, and 8 weeks, data from the initial and last assessment were used. Insufficient data were available to analyze follow-up period results. Treatment effects for crossover trials were calculated by using all assessments from the intervention and control periods and analyzing them as if they were a parallel group trial of the intervention versus control group (42). While this could result in a unit-of-analysis error, this is a conservative approach in which studies may result in being under versus over weighted (42). Given the primary outcomes and expected distribution of findings, this approach was believed to be superior to alternative approaches such as only including data from the first assessment period or attempting to impute standard deviations (42).

Pooled estimates for changes in outcomes—Separate *network (geometry) plots* for each outcome were used to provide a visual representation of the evidence base. While the *a priori* plan was to weight nodes (circles) by the number of participants randomized to each treatment and edges (lines) by the number of studies evaluating each pair of treatments, a *post hoc* decision was made to weight nodes and edges by the number of studies involved in each treatment and comparison, respectively (43). *Contribution plots* for each outcome were used to determine the most dominant comparisons for each network estimate as well as for the entire network (43). The weights applied were a function of the variance of the direct treatment effect and the network structure, the result being a percent contribution of each direct comparison to each network estimate (43). Network and contribution plots were generated using the *networkplot* and *netweight* commands, respectively (43), in STATA, version 14.1 (44).

Network meta-analysis was performed using a recently described *multivariate random-effects meta-regression model* (41), an approach that can be performed within a frequentist setting, allows for the inclusion of potential covariates, and correctly accounts for the

correlations from multi-arm trials (41). The frequentist approach was chosen over alternative Bayesian models because it avoids sensitivity to priors and Monte Carlo error (45). Non-overlapping 95% CIs were considered to represent statistically significant changes. Separate network meta-analysis models were used to examine for changes in BMI z-score and body weight. For those studies that reported data for changes in BMI z-score and body weight using both intention-to-treat and per protocol approaches, a post hoc decision was made to prioritize analyses by taking the more conservative approach of only including intention-to-treat results. However, analysis was also conducted by only including per protocol results.

Both *consistency* and *inconsistency* models were fit for changes in BMI z-score and body weight (41). Consistency models allow for intervention effects to be heterogeneous between studies but assumes no systematic differences between designs while inconsistency models allow for intervention effects to differ between designs to a greater degree than can be explained by the heterogeneity (41). For inconsistency, i.e., design by treatment model, the approach of Higgins et al., was used (46). This produces the Wald statistic as a global assessment for inconsistency. An alpha value 0.05 was considered to represent statistically significant inconsistency. Both consistency and inconsistency models were conducted using the recently updated mvmeta command for multivariate random-effects meta-regression (41) in STATA, version 14.1 (44).

Prediction intervals were used to enhance interpretation of results with respect to the magnitude of heterogeneity as well as provide an estimate of expected results in a future study (47,48). These were computed as $\bar{\mu} \pm t_{1-\alpha/2, df} \sqrt{\bar{v} + \tau^2}$, where $\bar{\mu}$ is the average weighted estimate across studies, $t_{1-\alpha/2, df}$ is the 100(1- $\alpha/2$) percentile of the t-distribution degrees of freedom, $\bar{v}$ is the estimated squared standard error of $\bar{\mu}$, and τ^2 is the estimated between study variance. For network meta-analysis, degrees of freedom (*df*) were set as the number of studies – the number of comparisons – 1 (49). Prediction intervals were produced using the mvmeta (41) and intervalplot (43) commands in STATA, version 14.1 (44).

Small-study-effects (publication bias, etc.) were assessed using comparison adjusted funnel plots (43,50). Unlike traditional funnel plots in pairwise meta-analysis, funnel plots in network meta-analysis need to account for the fact that studies estimate treatment effects for different comparisons. Consequently, there is no single reference line from which symmetry can be evaluated. For the comparison adjusted funnel plot, the horizontal axis represents the difference between study-specific effect sizes from the comparison-specific summary effect. In the absence of small-study effects, the comparison adjusted funnel plot should be symmetric around the zero line. Since the treatments need to be organized in some meaningful way to examine how small studies may differ from large ones, comparisons were defined so that all refer to an active treatment versus a control group. Comparison adjusted funnel plots were produced using the netfunnel command (43) in STATA, version 14.1 (44).

Transitivity, i.e., similarity in the distribution of potential effect modifiers across the different pairwise comparisons for each outcome (51), was assessed using random-effects network meta-regression analyses while controlling for different study designs within each comparison. Potential effect modifiers, determined a priori, included age, gender, baseline BMI z-score or body weight, length of training in weeks, and publication status (published

versus unpublished). Post hoc, gender and publication status were not examined because only 2 results were available for females and all studies were published. Transitivity analysis was conducted using the mvmeta command (41) in STATA, version 14.1 (44).

Ranking analysis, a major advantage of network meta-analysis, is the ability to rank all interventions for the outcome of interest. For this proposed project, ranking plots for a single outcome using probabilities were used (50,52). However, since the ranking of treatments based solely on the probability of each treatment being the best should be avoided because it does not account for the uncertainty in the relative treatment effects as well as the potential for assigning higher ranks for treatments in which little evidence is available, separate *rankograms and cumulative ranking probability plots* were used to present ranking probabilities along with their uncertainty for changes in BMI z-score and body weight (50,52). A rankogram for a specific treatment y is a plot of the probabilities of assuming each of the T ranks where T is the total number of treatments in the network (50,52). The cumulative rankograms display the probabilities that a treatment would be among the n best treatments, where n ranges from one to T (50,52). The surface under the cumulative ranking curve (SUCRA), a transformation of the mean rank, provides a hierarchy of treatments and accounts for the location and variance of all treatment effects (50,52). Larger SUCRA values are indicative of better ranks for the treatment (50,52). Separate ranking analyses for BMI z-score and body weight were conducted using the mvmeta(41) and SUCRA (43) commands in STATA, version 14.1 (44).

To improve practical application, a post hoc decision was made to estimate the number needed to treat (NNT) as well as percentile improvement for those comparisons in which statistically significant improvements in BMI z-score and body weight were observed. This was accomplished by first converting overall changes in BMI z-score and body weight into standardized mean difference effect sizes (53). Given the lack of information available on control group risk, the NNT was then estimated from the approach of Kraemer and Kupfer (54). Percentile improvement was estimated using Cohen's U_3 index.(55)

Software used for other analyses—Descriptive statistics were analyzed using Stata (version 14.1) (44), Microsoft Excel 2010 (36), and two add-ins for Excel, SSC-Stat (version 2.18) (56), and EZ-Analyze (version 3.0) (57). Cohen's kappa and U_3 statistics were calculated using Microsoft Excel 2010 (36).

Confidence in cumulative evidence—The strength of the body of evidence was assessed using a modification of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) instrument for network meta-analysis (58). This included evaluating the confidence in specific pairwise effect estimates as well as treatment rankings in a network meta-analysis across five domains: (1) study limitations, (2) joint consideration of indirectness and intransitivity, (3) joint consideration of statistical heterogeneity and statistical inconsistency, (4) imprecision and (5) publication bias (58). From these assessments, the overall strength of evidence was ranked as either high, moderate, low or very low (58). For randomized trials, scoring commenced at a high level and could be reduced by up to 2 levels for each domain and up to a maximum of 3 levels

across all 5 domains (58). All assessments were conducted following the same procedures as for the abstraction of data and risk of bias for individual studies.

Results

Characteristics of included studies

Of the 6,087 citations screened, 34 studies representing 73 groups (40 exercise, 33 control) and up to 2,239 participants (1,400 exercise, 839 control) met the criteria for inclusion (31,32,59-90). The number needed to screen was 0.56% while the NNR was 179. Reasons for exclusion, in order of prevalence, included inappropriate study design (47.7%), inappropriate intervention (17.8%), inappropriate population (17.1%), other reasons, for example, editorials (14.2%), inappropriate comparison (1.8%), inappropriate outcome (1.4%), and could not retrieve data (0.03%). A flow diagram that depicts the search process is shown in Figure 1 while a list of excluded studies, including the reasons for exclusion, can be found in Supplementary File 2. Of the 34 included studies, 29 (85.3%) were limited to 1 exercise and 1 control group that met all eligibility criteria (31,32,60-63,65-73,76-82,84-90), 3 (8.8%) included 2 exercise groups and 1 control group (64,74,75), 1 (2.9%) include three exercise groups and one control group (83), while another (2.9%) included 2 exercise groups that met our criteria (59). A total of 15 different requests for data were made to authors, 2 (13.3%) of which provided such.

All studies were published in 30 different journals between 2004 and 2015 ($\pm$ SD, 2010 $\pm$ 3.2, Mdn = 2011) (31,32,59-90), 31 (91.2%) in English-language journals (31,32,59,60,63-75,77-90), 2 (5.9%) in Chinese journals (62,76), and 1 (2.9%) in a Spanish-language journal (61). The studies were conducted in 15 different countries, 10 in the United States (32,64,71,74,75,79,80,82,87,89), 4 in China (62,76,85,86), 3 in South Korea (72,73,84), 2 each in either Australia (31,88), Brazil (59,61), Canada (60,83), Iran (67,68), or Turkey (70,81), and 1 each in Germany (78), New Zealand (77), Singapore (90), Sweden (69), Switzerland (66), Tunisia (65), or the United Kingdom (63). Twenty-eight studies used a non-intervention control group (31,32,60-62,65-82,84-88), 2 used an attention control group (63,89), while 3 employed alternative control groups (64,83,90) that included either (1) monthly lifestyle education classes on healthy diet, physical activity and stress management for all families (64), (2) dietary counseling and an energy deficit of 250 kcal per day for both intervention and control groups (83), or (3) participation by both exercise and control groups in physical education twice per week for 40 minutes per session (90). Five studies used some type of matching procedure prior to randomization (32,59,64,77,83). Two used race and sex (64,77), while 1 study each matched boys and girls according to either age, sex and BMI (32) sex and BMI (59), or sex and degree of overweight (83). Two studies used a cross-over design (31,88). Direct evidence was available from 4 (11.8%) of the included studies (59,74,75,83). For the 22 studies (64.7%) in which it could be determined, 10 used the per protocol approach to analyze their data (59,60,69,71,73,78,80,82,84,87), 6 used intention-to-treat (63,64,77,85,86,89), and 5 used both (61,66,74,75,83). One study reported no dropouts (76). Eleven of the 34 studies (32.4%) reported conducting sample size estimates prior to the start of their investigation (31,61,63,64,66,71,74,77,83,84,89). With regards to funding, 25 studies (73.5%) reported

receiving external and/or internal funding for their research from government, university and private sources (31,32,59,60,63-66,69,71-75,77-80,82-85,88-90).

Participant characteristics

A description of the physical characteristics of the participants is shown in Tables 1 and 2. Nineteen of the studies included both boys and girls (31,32,59-61,63,64,66,69,77-80,83,85-89), 13 were limited to boys (62,65,67,68,70-74,76,82,84,90) and 2 were limited to girls (75,81). For the 21 studies (61.8%) that reported information on race/ethnicity (31,59,62-64,71-77,80,82-87,89,90), participants, according to the original study authors, included Whites, Blacks, Hispanics, Other Hispanics, Asians, Maori, Pacific, Arabic and ZN/Euro. Nine studies reported that none of the participants smoked cigarettes (31,67,74-76,78,87,88,90) while 1 reported that none of the participants consumed alcohol (76). Three studies reported no change in physical activity beyond the actual exercise interventions in either the exercise or control groups (63,66,83) while 1 reported a decrease in both groups (69). For hyperlipidemia, 5 studies reported that none of the boys and girls were hyperlipidemic (32,67,68,76,88) while 2 reported that some were (66,80). Eighteen studies reported that none of the participants had type 1 (32,59,61,63,65,67,68,71,72,74-76,78,80,82-84,90) or type 2 diabetes (32,59,60,65,67,68,71,72,74-76,78,80,82-84,87,90) while 1 reported that some had pre-diabetes (64). With respect to hypertension, 9 studies reported that none of the boys and girls were hypertensive (32,67,68,76,80,83,87,88,90) while 2 reported that some were (66,82). Five studies reported that none of the participants had metabolic syndrome (65,67,68,76,83) while three reported the presence of such (32,79,82). One study reported that participants were insulin resistant (66) while another reported impaired flow-mediated dilation (79). For the 10 studies (29.4%) in which maturational stage was reported (32,59,60,64-66,82,83,85,87), 3 included those at the pre- and pubertal stage (32,64,85), 2 at either the pre-pubertal (60,66), pubertal and post-pubertal stages (82,87), or post-pubertal stages (59,83), and 1 at the pre-pubertal, pubertal and post-pubertal stages (65). Dropouts for the 18 studies (52.9%) that reported adequate information (32,61,63,64,66,71-73,75-78,82,83,85-87,89) ranged from 0% to 34% in the 23 exercise groups ($\pm$ SD, 12.5 ± 11.5 , Mdn = 10) and 0% to 33% in the 18 control groups ($\pm$ SD, 12.5 ± 11.3 , Mdn = 15). Reasons for dropping out included work, changes in school hours, lack of funds for transportation, lack of motivation, unhappy with group assignment, refusal to complete testing, moved, medical condition that precluded continued participation, time, scheduling conflicts and pregnancy.

Exercise program characteristics

Pooled exercise intervention characteristics are shown in Table 3 while a study level description is provided in Supplementary File 3. Of the 40 exercise groups, 25 (62.5%) participated in aerobic types of exercise (31,32,59,61,63-65,68,70,72,74-81,83-87,89), 5 (12.5%) in strength training (71,74,75,82,83) and 11 (25.0%) in both (59,60,62,66,67,69,73,83,88,90). Aerobic activities included walking, jogging, cycling, swimming, games, including active video games, stair climbing, jumping rope, dancing, team sports and gymnastics (31,32,59,59-70,72-81,83,83-90). For those studies that included strength training, the within-study number of sets ranged from 1 to 4 for the 9 groups

(56.3%) that reported such information (66,67,71,74,75,82,83,90), the number of repetitions from 3 to 25 for the 11 groups (68.8%) reporting (59,60,66,67,71,74,75,82,83,90), and the rest period between sets from 45 to 180 seconds for the 5 groups (31.3%) reporting information (59,67,74,82,90). The number of strength training exercises performed for the 9 groups (56.3%) reporting data ranged from 3 to 13 ($\pm$ SD, 9 ± 3 , Mdn = 10) (59,60,67,73-75,82,83).

For the 24 of 40 groups (60.0%) in which data were available, 11 participated in moderate to vigorous intensity exercise (31,32,60,66,69,73-75,83,84,88), 4 in moderate intensity exercise (63,74,75,81), 3 in either light to moderate (76,84,85) or vigorous (64,86,87) intensity exercise, 2 in very light, light, moderate and vigorous intensity exercise (67,90) and one in moderate, vigorous and near-maximal exercise (82). With respect to the delivery of exercise, 32 groups (80.0%) participated in supervised exercise (31,32,59-62,64-70,72-76,78,81,82,84-90), 4 in (10.0%) unsupervised exercise (71,77,79,80), and 4 (10.0%), of which 3 were from one study, participated in both (63,83).

Risk of bias assessment

A summary of risk of bias assessment findings is shown in Figure 2 while study-level results are shown in Supplementary File 4. As can be seen, all of the studies were considered to be at a high risk of bias with respect to the blinding of participants and personnel given the difficulty in blinding participants to exercise interventions (31,32,59-90). In addition, 91% (31,32,59-65,67-83,85,86,88-90) and 76% (31,32,59,61-63,65,67-70,72,73,76,78-82,84-90) of the studies, respectively, were considered to be at either a high or unclear risk of bias for blinding of outcome assessors and allocation concealment. Furthermore, more than half of the studies were classified as being at either a high or unclear risk of bias for boys and girls being physically inactive prior to enrollment (31,32,59-63,65,69,71,73,76,77,79-82,85-90) as well as incomplete outcome reporting (31,32,60-62,65,69,71,73,76,77,79-82,85,86,88,90). Finally, half the studies were considered to be at either a high or unclear risk for incomplete outcome data (31,32,59,60,62,65,68-70,72,73,78,79,81,82,84,88) while all but one of the studies (86) was considered to be at a low risk of bias for random sequence generation.

Data synthesis

Changes in BMI z-score—Changes in BMI z-score were available from 34 studies (31,32,59-90). Figure 3 shows a network plot of results for BMI z-score. As can be seen, the most common comparator was the control group followed by the aerobic, aerobic and strength, and strength training groups. The most common pairwise comparisons were aerobic versus control followed by aerobic and strength versus control, strength versus control, aerobic versus strength, aerobic versus aerobic and strength, and aerobic and strength versus strength. The contribution plot in Figure 4 demonstrates that the control versus aerobic exercise comparison provided the largest contribution to the entire network (29.4%).

Consistency and inconsistency results are shown in Figures 5 and 6. When limited to intention-to-treat findings for the three studies that also included data according to per

protocol results (74,75,83), statistically significant decreases in BMI *z*-score were observed for aerobic exercise as well as combined aerobic exercise and strength training. The NNT was 2 for both aerobic exercise and combined aerobic exercise and strength training while percentile improvements were 28.8% for aerobic exercise and 31.5% for combined aerobic exercise and strength training. Overlapping 95% PIs were observed for both, suggesting that the comparisons may be affected by common heterogeneity in the network (Figure 5). No other statistically significant changes were observed for any of the other comparisons. Visual inspection of Figure 6 shows the similarity of “pooled within design” and “pooled overall” results. These findings suggest no evidence of inconsistency between the different designs within each treatment comparison. In addition, the global assessment for inconsistency was not statistically significant. Thus, both results were suggestive of consistency. Funnel plot asymmetry for potential small study effects was observed for all three interventions when compared to controls (Figure 7). When examined for *transitivity*, none of the potential effect modifiers (age, baseline BMI *z*-score, length of training in weeks) were statistically significant predictors of changes in BMI *z*-score for any of the treatments (results not shown).

Surface under the cumulative ranking curve (SUCRA) results for changes in BMI *z*-score are shown in Table 4 and Figure 8. As can be seen, the best treatment was combined aerobic and strength training followed by aerobic exercise, strength training, and control. Overall, estimated and predictive probability results according to order were similar (Figure 8). Results for BMI *z*-score were similar when pooling was limited to per protocol findings for the three studies that also included intention-to-treat findings (results not shown) (74,75,83).

Strength of the evidence for BMI *z*-score treatment effect comparisons based on a modification of the GRADE instrument specific to network meta-analysis is shown in Table 5. As can be seen, confidence in the results were considered low for 2 treatments and very low for the other 5. All of the treatments were downgraded based on study limitations and inconsistency, 4 for imprecision and 1 for potential publication bias. More detailed information for the basis of these judgements is shown in Supplementary Files 5 and 6. *Changes in body weight.* Changes in body weight, a secondary outcome, were available from 28 studies (31,32,59-62,65-77,79-85,88,90). Supplementary File 7 shows a network plot of results for body weight. Similar to BMI *z*-score, the most common comparator was the control group followed by the aerobic, aerobic and strength, and strength training groups. The most common pairwise comparisons were aerobic versus control followed by aerobic and strength versus control, strength versus control, aerobic versus strength, aerobic versus aerobic and strength, and aerobic and strength versus strength. An examination of the contribution plot in Supplementary File 8 demonstrates that the control versus aerobic exercise comparison provided the largest contribution to the entire network (30.7%).

Consistency and inconsistency results are shown in Supplementary Files 9 and 10. When limited to intention-to-treat findings for the three studies that also included data according to per protocol results (74,75,83), statistically significant decreases in body weight were observed for aerobic exercise as well as combined aerobic exercise and strength training. The NNT was 3 for aerobic exercise and 2 for combined aerobic exercise and strength training while percentile improvements were 25.4% for aerobic exercise and 27.3% for

combined aerobic exercise and strength training. Overlapping 95% PIs were observed for both, suggesting that the comparisons may be affected by common heterogeneity in the network (Supplementary File 9). No other statistically significant changes were observed for any of the other comparisons. Visual inspection of Supplementary File 10 shows the similarity of “pooled within design” and “pooled overall” results. These findings suggest no evidence of inconsistency between the different designs within each treatment comparison. In addition, the global assessment for inconsistency was not statistically significant. Thus, both results are suggestive of consistency. Funnel plot asymmetry for potential small study effects was observed (Supplementary File 11). When examined for *transitivity*, none of the potential effect modifiers that could be assessed (age, baseline body weight, length of training in weeks) were statistically significant predictors of changes in body weight for any of the treatments (results not shown).

Surface under the cumulative ranking curve (SUCRA) results for changes in body weight are shown in Supplementary Files 12 and 13. As can be seen, the best treatment was combined aerobic and strength training followed by aerobic exercise only, strength training, and control. Overall, estimated and predictive probability results with respect to order were similar (Supplementary File 13). Results for BMI *z*-score were similar when pooling was limited to per protocol findings for the three studies that also included intention-to-treat findings (results not shown) (74,75,83).

Overall strength of evidence results for treatment effect comparisons for body weight are shown in Supplementary File 14. As can be seen, results were the same as for BMI *z*-score; confidence in the results were considered low for 2 treatments and very low for the other 5. All treatments were downgraded based on study limitations and inconsistency as well as 4 for imprecision and 1 for potential publication bias. More detailed information for the basis of these judgements is shown in Supplementary Files 15 and 16.

Discussion

Overall Findings

Based on non-overlapping confidence intervals, the overall findings of the current network meta-analysis suggest that aerobic exercise as well as aerobic exercise combined with strength training reduce BMI *z*-score in overweight and obese children and adolescents. However, findings for strength training were inconsistent, suggesting that some may benefit while others may not. Confidence in these findings are reinforced by a lack of inconsistency and apparent intransitivity. With respect to which intervention is best when limited to statistically significant findings only, combined aerobic exercise and strength training was consistently ranked as the best treatment followed by aerobic exercise. However, all prediction intervals included zero (0), suggesting that results for a future study may vary from beneficial to non-beneficial. In addition, while no statistically significant change in BMI *z*-score was observed for strength training, this may have been confounded by increases in lean muscle mass. Findings similar to the above were observed for body weight.

The results of the current study compare favorably to a previous aggregate data meta-analysis by the investigative team (19). Across all types of exercise (aerobic, strength

training, both) BMI z-score was significantly reduced (, -0.06, 95% CI, -0.09 to -0.03) (19). The statistically significant reductions in the current study, specific to type of exercise, are approximately 0.04 greater (40%) for aerobic exercise and 0.05 larger (45%) for combined aerobic and strength training. Thus, while the relative differences are large, the absolute values are not.

Implications for research

The results of the current study have several implications for research. With respect to future meta-analytic research, BMI z-score was chosen as the primary outcome in the current meta-analysis because it has been reported to be more valid than other BMI measures in children and adolescents (91). However, BMI in kg-m² continues to be the most frequently assessed and reported measure of adiposity in both the clinical and public health setting. Given the former, it would appear plausible to suggest that a need exists for a network meta-analysis addressing the effects of selected types of exercise (aerobic, strength, both) on BMI in kg-m² among overweight and obese children and adolescents. In addition, since all types of BMI measures as well as body weight do not capture changes in body composition (fat mass, fat-free mass, etc.), the inclusion of such outcomes is also recommended. Such analyses might provide one with more accurate information regarding the effects of different types of exercise on selected measures of body composition in overweight and obese children and adolescents.

With respect to the conduct of future randomized trials, the small number, as well as equivocal results available for strength training as a sole intervention, suggests that further work in this area may be necessary. In addition, it appears that a need exists for studies that examine the dose-response effects of exercise on measures of adiposity, including BMI z-score, in overweight and obese children and adolescents. Furthermore, the absence of an association in the current study between changes in BMI z-score with age, initial BMI z-score and length of training in weeks does not necessarily mean that such an association does not exist. Thus, it is suggested that these potential covariates be examined in original randomized trials. Along those lines, the inability to examine the association between gender and changes in BMI z-score in the current investigation warrants such analysis in original randomized trials. Finally, the monitoring of physical activity levels beyond the intervention for all participants during the study is recommended given the potential for physical activity compensation (92). As an example, one of the studies included in the current network meta-analysis found that when compared to controls, total daily physical activity decreased in the exercise group (69).

The results of the current systematic review with meta-analysis suggest several areas for improvement in the reporting of randomized trials addressing the effects of exercise on adiposity in overweight and obese children and adolescents. First, based on results using the Cochrane Risk of Bias Instrument, it is suggested that future randomized trials do a better job of reporting information on (1) allocation concealment, (2) blinding of outcome assessment, (3) incomplete outcome reporting and (4) the physical activity levels of the participants prior to study enrollment. While all studies were considered to be at a high risk of bias for blinding of participants and personnel, it is important to understand that it is

virtually impossible to blind participants in an exercise intervention trial. Thus, this potential risk derives from the nature of the intervention versus any weaknesses in an investigative team's study design. Second, there was a lack of information reported on adverse events and the cost-effectiveness of the exercise interventions employed. Since adverse events and cost-effectiveness are important determinants in the decision of what interventions to recommend for whom, the reporting of such information in future trials is recommended. Third, it is suggested that future randomized trials report information on compliance of the participants to the exercise intervention in which they are assigned. In the current meta-analysis, data on compliance was available for only 32.5% of the exercise intervention groups.

Implications for practice

While additional high-quality research is needed regarding the effects of different types of exercise on BMI *z*-score in overweight and obese children and adolescents before any type of certainty might be established, the findings of the current study suggest that aerobic exercise as well as combined aerobic exercise and strength training may lower BMI *z*-score as well as body weight. In addition, the low NNT (2 to 3) and percentile improvements (25.4% to 31.5%) for BMI *z*-score as well as body weight lend support for the promotion of these types of interventions in both the clinical and public health setting. Given these findings, as well as the fact that exercise is probably a safe activity for most children and adolescents, it would seem plausible to suggest that overweight and obese children and adolescents participate in both aerobic exercise and strength training, especially given the numerous other benefits that might be derived from participation in such (93). With the following in mind, adherence to recommendations from the Centers for Disease Control and Prevention (93) as well as the World Health Organization would seem appropriate (94). Broadly, this includes at least 60 minutes per day of moderate to vigorous intensity physical activity with an emphasis on aerobic types of exercise (walking briskly, bicycling, swimming, etc.) as well as strengthening exercises at least 3 days per week. While beyond the scope of the current study, it would also appear plausible to suggest that a combination of aerobic exercise and strength training along with an energy restricted diet may be optimal for reducing BMI *z*-score as well as body weight and other measures of adiposity in overweight and obese children and adolescents.

Strengths and limitations

The results of the current study have at least three strengths. First, to the best of the authors' knowledge, this is the first systematic review with network meta-analysis to focus on the different types of exercise (aerobic, strength, both) and their effects on changes in BMI *z*-score in overweight and obese children and adolescents. This is important because the current approach allowed for the inclusion of both direct and indirect evidence in the analysis, thereby leading to more informed decisions. Second, the current findings provide useful and recent information to practitioners, researchers, and policy-makers with respect to the effects of exercise on changes in BMI *z*-score among overweight and obese children and adolescents. Third, the inclusion of PIs in the current study can be beneficial for investigators planning future randomized trials in which BMI *z*-score is a primary outcome.

In addition to strengths, the current study has at least six potential limitations. First, with the exception of aerobic exercise as well as combined aerobic and strength training, the number of effect sizes was limited to 5 or less for all other comparisons. Thus, it may be that the lack of a statistically significant improvement for some comparisons was due to the small number of effect sizes available for analysis. Second, the varying components nested within each type of exercise (intensity, duration, compliance, etc.) could have impacted the current findings. A third potential limitation, inherent in all aggregate data meta-analyses, is ecological fallacy. Thus, it may be that our group findings may not apply at the individual participant level. Fourth, like any aggregate data meta-analysis, the covariate results reported do not support causal inferences given that studies are not randomly assigned to covariates (95). Thus, such associations would need to be tested in adequately powered randomized trials. Fifth, a large number of statistical tests were conducted but no adjustments were made for such, thereby leading to the possibility of chance findings. However, no adjustments for multiple tests were made because of concerns about missing possibly important findings (96). Sixth, based on the GRADE assessment, the strength of evidence was rated as low or very low for the different comparisons. Thus, this may weaken the strength of the current findings. However, it is important to realize that while somewhat flexible, the GRADE assessment instrument may be overly conservative, thereby rendering it very difficult to derive an evidence rating of “high” (58).

Conclusions

Aerobic exercise and combined aerobic exercise and strength training are associated with reductions in BMI z-score. However, a need exists for additional, well designed randomized trials before certainty is determined.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Ng M, Fleming T, Robinson M, Thomson B, Graetz N, Margono C, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 2014; 384:766–81.
2. Fryar, CD., Carroll, MD., Ogden, CL. Prevalence of overweight and obesity among children and adolescents: United States, 1963-1965 through 2011-2012. *Health E-Stat* : 1-6. 2014. National Center for Health Statistics; Dec 2. 2015
3. Finkelstein EA, Graham WCK, Malhotra R. Lifetime direct medical costs of childhood obesity. *Pediatrics*. 2014; 133:854–62. [PubMed: 24709935]
4. Centers for Disease Control and Prevention. Childhood obesity facts. Atlanta, Georgia: U.S. Department of Health & Human Services; Aug 27. 2015 12-3-2015
5. Freedman DS, Mei Z, Srinivasan SR, Berenson GS, Dietz WH. Cardiovascular risk factors and excess adiposity among overweight children and adolescents: the Bogalusa Heart Study. *J Pediatr*. 2007; 150:12–7. [PubMed: 17188605]
6. Li C, Ford ES, Zhao G, Mokdad AH. Prevalence of pre-diabetes and its association with clustering of cardiometabolic risk factors and hyperinsulinemia among U.S. adolescents: National Health and Nutrition Examination Survey 2005-2006. *Diabetes Care*. 2009; 32:342–7. [PubMed: 18957533]

7. Daniels SR, Arnett DK, Eckel RH, Gidding SS, Hayman LL, Kumanyika S, et al. Overweight in children and adolescents: pathophysiology, consequences, prevention, and treatment. *Circulation*. 2005; 111:1999–12. [PubMed: 15837955]
8. Dietz WH. Overweight in childhood and adolescence. *N Engl J Med*. 2004; 350:855–7. [PubMed: 14985480]
9. Guo SS, Chumlea WC. Tracking of body mass index in children in relation to overweight in adulthood. *Am J Clin Nutr*. 1999; 70:145S–8S.
10. Freedman DS, Khan LK, Serdula MK, Dietz WH, Srinivasan SR, Berenson GS. The Relation of Childhood BMI to Adult Adiposity: The Bogalusa Heart Study. *Pediatrics*. 2005; 115:22–7. [PubMed: 15629977]
11. Freedman DS, Wang J, Thornton JC, Mei Z, Sopher AB, Pierson RN Jr, et al. Classification of body fatness by body mass index-for-age categories among children. *Arch Pediatr Adolesc Med*. 2009; 163:805–11. [PubMed: 19736333]
12. Freedman DS, Khan LK, Dietz WH, Srinivasan SR, Berenson GS. Relationship of childhood obesity to coronary heart disease risk factors in adulthood: the Bogalusa Heart Study. *Pediatrics*. 2001; 108:712–8. [PubMed: 11533341]
13. Schranz N, Tomkinson G, Olds T. What is the effect of resistance training on the strength, body composition and psychosocial status of overweight and obese children and adolescents? A systematic review and meta-analysis *Sports Med*. 2013; 43:893–907. [PubMed: 23729196]
14. Atlantis E, Barnes EH, Singh MA. Efficacy of exercise for treating overweight in children and adolescents: a systematic review. *Int J Obes (Lond)*. 2006; 30:1027–40. [PubMed: 16534526]
15. Harris KC, Kuramoto LK, Schulzer M, Retallack JE. Effect of school-based physical activity interventions on body mass index in children: a meta-analysis. *Can Med Assoc J*. 2009; 180:719–26. [PubMed: 19332753]
16. McGovern L, Johnson JN, Paulo R, Hettinger A, Singhal V, Kamath C, et al. Treatment of Pediatric Obesity: A Systematic Review and Meta-Analysis of Randomized Trials. *J Clin Endocrinol Metab*. 2008; 93:4600–5. [PubMed: 18782881]
17. Cesa CC, Sbruzzi G, Ribeiro RA, Barbiero SM, de Oliveira Petkowicz R, Eibel B, et al. Physical activity and cardiovascular risk factors in children: meta-analysis of randomized clinical trials. *Prev Med*. 2014; 69:54–62. [PubMed: 25175591]
18. Guerra PH, Nobre MRC, da Silveira JAC, Taddei JADC. The effect of school-based physical activity interventions on body mass index: a meta-analysis of randomized trials. *Clinics*. 2013; 68:1263–73. [PubMed: 24141844]
19. Kelley GA, Kelley KS, Pate RR. Effects of exercise on BMI z-score in overweight and obese children and adolescents: a systematic review with meta-analysis. *BMC Pediatrics*. 2014; 14:225. [PubMed: 25204857]
20. Kelley GA, Kelley KS, Pate RR. Exercise and BMI in overweight and obese children and adolescents: A systematic review with trial sequential meta-analysis. *Biomed Res Int*. 2015; 2015:1–17.
21. Li T, Puhan MA, Vedula SS, Singh S, Dickersin K. Network meta-analysis-highly attractive but more methodological research is needed. *BMC Med*. 2011; 9:79. [PubMed: 21707969]
22. Hutton B, Salanti G, Caldwell DM, et al. The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. *Ann Int Med*. 2015; 162:777–84. [PubMed: 26030634]
23. Kelley GA, Kelley KS. Exercise and BMI z-score in overweight and obese children and adolescents: protocol for a systematic review and network meta-analysis of randomised trials. *BMJ Open*. 2016; 6:1–7.
24. Physical Activity Guidelines Advisory Committee. Physical Activity Guidelines Advisory Report. Washington, DC: U.S. Department of Health and Human Services; 2008.
25. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985; 100:126–31. [PubMed: 3920711]
26. Sacks HS, Chalmers TC, Smith H. Randomized versus historical controls for clinical trials. *Am J Med*. 1982; 72:233–40. [PubMed: 7058834]

27. Schulz KF, Chalmers I, Hayes R, Altman DG. Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995; 273:408–12. [PubMed: 7823387]
28. Jago R, Jonker ML, Missaghian M, Baranowski T. Effect of 4 weeks of Pilates on the body composition of young girls. *Prev Med*. 2006; 42:177–80. [PubMed: 16376979]
29. Dietz WH. Health consequences of obesity in youth: childhood predictors of adult disease. *Pediatrics*. 1998; 101:518–25. [PubMed: 12224658]
30. Watts K, Beye P, Siafarikas A, Davis EA, Jones TW, O'Driscoll G, et al. Exercise training normalizes vascular dysfunction and improves central adiposity in obese adolescents. *J Am Coll Cardiol*. 2004; 43:1823–7. [PubMed: 15145107]
31. Watts K, Beye P, Siafarikas A, et al. Effects of exercise training on vascular function in obese children. *J Pediatr*. 2004; 144:620–5. [PubMed: 15126996]
32. Kelly AS, Wetzsteon RJ, Kaiser DR, Steinberger J, Bank AJ, Dengel DR. Inflammation, insulin, and endothelial function in overweight children and adolescents: The role of exercise. *J Pediatr*. 2004; 145:731–6. [PubMed: 15580192]
33. Kelly AS. The effects of aerobic exercise training on vascular structure and function in obese children. *DAI*. 2004; 65:185.
34. Version 12.0.3. Philadelphia, PA: Thompson ResearchSoft; 2009. Reference Manager [computer program].
35. Lee E, Dobbins M, DeCorby K, Mcrae L, Tirilis D, Husson H. An optimal search filter for retrieving systematic reviews and meta-analyses. *BMC Med Res Methodol*. 2012; 12:51. [PubMed: 22512835]
36. Version 2010. Redmond, Washington: Microsoft Corporation; 2010. Microsoft Excel [computer program].
37. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968; 70:213–20. [PubMed: 19673146]
38. Higgins JP, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011; 343:d5928. [PubMed: 22008217]
39. Ahn S, Becker BJ. Incorporating quality scores in meta-analysis. *J Educ Behav Stat*. 2011; 36:555–85.
40. Follmann D, Elliot P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. *J Clin Epidemiol*. 1992; 45:769–73. [PubMed: 1619456]
41. White IR. Network meta-analysis. *Stata J*. 2015; 15:951–85.
42. Higgins JPT, Green, S. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011. The Cochrane Collaboration; 2011.
43. Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: The network graphs package. *Stata J*. 2015; 15:905–50.
44. Version 14.1. College Station, TX: StataCorp LP; 2015. Stata Statistical Software: Release 14 [computer program].
45. White IR, Barrett JK, Jackson D, Higgins JPT. Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression. *Res Syn Meth*. 2012; 3:111–25.
46. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Syn Meth*. 2012; 3:98–110.
47. Higgins JP, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. *J R Stat Soc Ser A*. 2009; 172:137–59.
48. Kelley GA, Kelley KS. Impact of progressive resistance training on lipids and lipoproteins in adults: Another look at a meta-analysis using prediction intervals. *Prev Med*. 2009; 49:473–5. [PubMed: 19804794]
49. Cooper, HC., Hedges, LV., Valentine, JF. *The handbook of research synthesis*. New York, New York: Russell Sage; 2009.
50. Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS ONE*. 2013; 8

51. Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med.* 2013; 11
52. Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol.* 2011; 64:163–71. [PubMed: 20688472]
53. Hedges, LV., Olkin, I. Statistical methods for meta-analysis. San Diego, CA: Academic Press; 1985.
54. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry.* 2006; 59:990–6. [PubMed: 16368078]
55. Cohen, J. Statistical power analysis for the behavioral sciences. New York: Academic Press; 1988.
56. Version 2.18. University of Reading, United Kingdom: Statistical Services Center; 2007. SSC-Stat [computer program].
57. EZ Analyze [computer program]. Version 3.0. 2007
58. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. *PLoS ONE.* 2014; 9
59. Ackel-D'Elia C, Carnier J, Bueno CR Jr, et al. Effects of different physical exercises on leptin concentration in obese adolescents. *Int J Sports Med.* 2014; 35:164–71. [PubMed: 23868679]
60. Alberga AS, Farnesi BC, Lafleche A, Legault L, Komorowski J. The effects of resistance exercise training on body composition and strength in obese prepubertal children. *Phys Sportsmed.* 2013; 41:103–9. [PubMed: 24113708]
61. Alves JG, Gale CR, Souza E, Batty GD. Effect of physical exercise on bodyweight in overweight children: a randomized controlled trial in a Brazilian slum. *Cad Saude Publica.* 2008; 24:S353–S359. [PubMed: 18670715]
62. Cheng HL, Peng P, Zhu R, Qin YS. Effects of eight weeks exercise prescription intervention on aerobic capacity, body composition, blood lipid and C-reactive protein in obese adolescents. [Chinese]. *J Jilin Univ Med Ed.* 2012; 38:745–9.
63. Daley AJ, Copeland RJ, Wright NP, Roalfe A, Wales JK. Exercise therapy as a treatment for psychopathologic conditions in obese and morbidly obese adolescents: a randomized, controlled trial. *Pediatrics.* 2006; 118:2126–34. [PubMed: 17079587]
64. Davis CL, Pollock NK, Waller JL, et al. Exercise dose and diabetes risk in overweight and obese children: a randomized controlled trial. *JAMA.* 2012; 308:1103–12. [PubMed: 22990269]
65. Elloumi M, Makni E, Ounis OB, et al. Six-minute walking test and the assessment of cardiorespiratory responses during weight-loss programmes in obese children. *Physiother Res Int.* 2011; 16:32–42. [PubMed: 21110411]
66. Farpour-Lambert NJ, Aggoun Y, Marchand LM, Martin XE, Herrmann FR, Beghetti M. Physical activity reduces systemic blood pressure and improves early markers of atherosclerosis in pre-pubertal obese children. *J Am Coll Cardiol.* 2009; 54:2396–406. [PubMed: 20082930]
67. Fazelifar S, Ebrahim K, Sarkisian V. Effect of concurrent training and detraining on anti-inflammatory biomarker and physical fitness levels in obese children. *Rev Bras Med Esporte.* 2013; 19:349–54.
68. Ghorbanian B, Ravassi A, Kordi MR, Hedayati M. The effects of rope training on lymphocyte ABCA1 expression, plasma ApoA-I and HDL-c in boy adolescents. *Int J Endocrinol Metab.* 2013; 11:76–81. [PubMed: 23825977]
69. Hagstromer M, Elmberg K, Marild S, Sjostrom M. Participation in organized weekly physical exercise in obese adolescents reduced daily physical activity. *Acta Paediatr.* 2009; 98:352–4. [PubMed: 19006526]
70. Karacabey K. The effect of exercise on leptin, insulin, cortisol and lipid profiles in obese children. *J Int Med Res.* 2009; 37:1472–8. [PubMed: 19930853]
71. Kelly LA, Loza A, Lin X, et al. The effect of a home-based strength training program on type 2 diabetes risk in obese Latino boys. *J Pediatr Endocrinol Metab.* 2015; 28:315–22. [PubMed: 25781537]
72. Kim ES, Im JA, Kim KC, et al. Improved insulin sensitivity and adiponectin level after exercise training in obese Korean youth. *Obesity (Silver Spring).* 2007; 15:3023–30. [PubMed: 18198311]

73. Kim HJ, Lee S, Kim TW, et al. Effects of exercise-induced weight loss on acylated and unacylated ghrelin in overweight children. *Clin Endocrinol (Oxf)*. 2008; 68:416–22. [PubMed: 17888020]
74. Lee S, Bacha F, Hannon T, Kuk JL, Boesch C, Arslanian S. Effects of aerobic versus resistance exercise without caloric restriction on abdominal fat, intrahepatic lipid, and insulin sensitivity in obese adolescent boys: a randomized, controlled trial. *Diabetes*. 2012; 61:2787–95. [PubMed: 22751691]
75. Lee S, Deldin AR, White D, et al. Aerobic exercise but not resistance exercise reduces intrahepatic lipid content and visceral fat and improves insulin sensitivity in obese adolescent girls: a randomized controlled trial. *Am J Physiol Endocrinol Metab*. 2013; 305:E1222–E1229. [PubMed: 24045865]
76. Li GX, Wang ZY, Lyu W, Zhao LJ. Influence of aerobic exercise in heart rate variability in obese adolescents. [Chinese]. *J Jilin Univ Med Ed*. 2014; 40:1093–7.
77. Maddison R, Foley L, Ni MC, et al. Effects of active video games on body composition: a randomized controlled trial. *Am J Clin Nutr*. 2011; 94:156–63. [PubMed: 21562081]
78. Meyer AA, Kundt G, Lenschow U, Schuff-Werner P, Kienast W. Improvement of early vascular changes and cardiovascular risk factors in obese children after a six-month exercise program. *J Am Coll Cardiol*. 2006; 48:1865–70. [PubMed: 17084264]
79. Murphy EC, Carson L, Neal W, Baylis C, Donley D, Yeater R. Effects of an exercise intervention using Dance Dance Revolution on endothelial function and other risk factors in overweight children. *Int J Pediatr Obes*. 2009; 4:205–14. [PubMed: 19922034]
80. Rooney BL, Gritt LR, Havens SJ, Mathiason MA, Clough EA. Growing healthy families: family use of pedometers to increase physical activity and slow the rate of obesity. *Wis Med J*. 2005; 104:54–60.
81. Saygin O, Ozturk MA. The effect of twelve week aerobic exercise programme on health related physical fitness components and blood lipids in obese girls. *Afr J Pharm Pharmacol*. 2011; 5:1441–5.
82. Shaibi GQ, Cruz ML, Ball GD, et al. Effects of resistance training on insulin sensitivity in overweight Latino adolescent males. *Med Sci Sports Exerc*. 2006; 38:1208–15. [PubMed: 16826016]
83. Sigal RJ, Alberga AS, Goldfield GS, et al. Effects of aerobic training, resistance training, or both on percentage body fat and cardiometabolic risk markers in obese adolescents: the healthy eating aerobic and resistance training in youth randomized clinical trial. *Jama Pediatrics*. 2014; 168:1006–14. [PubMed: 25243536]
84. Song JK, Stebbins CL, Kim TK, Kim HB, Kang HJ, Chai JH. Effects of 12 weeks of aerobic exercise on body composition and vascular compliance in obese boys. *J Sports Med Phys Fitness*. 2012; 52:522–9. [PubMed: 22976739]
85. Sun MX, Huang XQ, Yan Y, et al. One-hour after-school exercise ameliorates central adiposity and lipids in overweight Chinese adolescents: a randomized controlled trial. *Chin Med J (Engl)*. 2011; 124:323–9. [PubMed: 21362327]
86. Tan S, Yang C, Wang J. Physical training of 9- to 10-year-old children with obesity to lactate threshold intensity. *Pediatr Exerc Sci*. 2010; 22:477–85. [PubMed: 20814042]
87. Wagener TL, Fedele DA, Mignogna MR, Hester CN, Gillasp SR. Psychological effects of dance-based group exergaming in obese adolescents. *Pediatr Obes*. 2012; 7:e68–e74. [PubMed: 22767495]
88. Watts K, Beye P, Siafarikas A, et al. Exercise training normalizes vascular dysfunction and improves central adiposity in obese adolescents. *J Am Coll Cardiol*. 2004; 43:1823–7. [PubMed: 15145107]
89. Weintraub DL, Tirumalai EC, Haydel KF, Fujimoto M, Fulton JE, Robinson TN. Team sports for overweight children: the Stanford Sports to Prevent Obesity Randomized Trial (SPORT). *Arch Pediatr Adolesc Med*. 2008; 162:232–7. [PubMed: 18316660]
90. Wong PC, Chia MY, Tsou IY, et al. Effects of a 12-week exercise training programme on aerobic fitness, body composition, blood lipids and C-reactive protein in adolescents with obesity. *Ann Acad Med Singapore*. 2008; 37:286–93. [PubMed: 18461212]

91. Inokuchi M, Matsuo N, Takayama JI, Hasegawa T. BMI z-score is the optimal measure of annual adiposity change in elementary school children. *Ann Hum Biol.* 2011; 38:747–51. [PubMed: 22014004]
92. Gomersall SR, Rowlands AV, English C, Maher C, Olds TS. The ActivityStat hypothesis: the concept, the evidence and the methodologies. *Sports Med.* 2013; 43:135–49. [PubMed: 23329607]
93. Centers for Disease Control and Prevention. How much physical activity do children need?. Atlanta, Georgia: 2013 Jun 24. 2013
94. World Health Organization. Physical activity and young people. Geneva, Switzerland: World Health Organization; 2016 Jun 10. 2016
95. Littell, JH., Corcoran, J., Pillai, V. Systematic reviews and meta-analysis. New York: Oxford University Press; 2008.
96. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiol.* 1990; 1:43–6.

Figure 1.

Flow diagram for selection of articles. *, number of reasons exceeds the number of articles because some articles were excluded for more than one reason.

Figure 2.
Risk of Bias. Risk of bias based on Cochrane's Risk of Bias Assessment Instrument.

Figure 3. Network plot for study comparisons included in BMI z-score analysis. The nodes (circles) represent the different treatments while the edges (lines) represent the available direct comparisons between pairs of treatments. Both nodes and edges are weighted by the number of studies involved in each treatment and comparison, respectively.

		Direct comparisons in the network					
		BvsA	CvsA	DvsA	BvsC	BvsD	CvsD
Network meta-analysis estimates	Mixed estimates						
	BvsA	67.4	9.0	5.3	13.0	1.2	4.1
	CvsA	28.0	22.2	10.8	28.2	0.2	10.6
	DvsA	8.0	5.2	67.5	6.0	2.0	11.3
	BvsC	24.6	17.1	7.5	41.7	0.7	8.3
	BvsD	36.2	2.2	38.4	11.7	2.0	9.4
	CvsD	17.6	14.6	32.2	18.9	1.4	15.3
Indirect estimates							
Entire network		29.4	11.8	27.2	20.1	1.3	10.2
Included studies		24	5	9	3	2	1

Figure 4.

Contribution plot for BMI z-score comparisons. The size of the squares are proportional to the percent contribution of the column-defining direct comparison to the row-defining network estimate. A = Control, B = Aerobic, C = Strength, D = Aerobic and Strength. Included Studies, number of observations (44) nested within studies (34).

Figure 5.

Interval plot for changes in BMI z-score based on all pairwise comparisons. The black horizontal lines represent the confidence intervals while the red lines represent the prediction intervals. The number of observations for each comparison were 24 (Aerobic versus Control), 5 (Strength versus Control), 9 (Aerobic and Strength versus Control), 3 (Strength versus Aerobic), 2 (Aerobic and Strength versus Aerobic) and 1 (Aerobic and Strength versus Strength).

Test of consistency: $\chi^2(6)=6.83$, $P=0.337$

Figure 6.

Network forest plot for changes in BMI z-score based on individual study results, grouped by treatment contrast and design. Markers for each point estimate are proportional to the inverse square of the standard error. Individual study results are shown as blue point estimates (squares) and 95% confidence intervals (lines), pooled estimates and 95% CIs within designs as hollow green diamonds, and overall pooled results and 95% CIs for each of the six treatment contrasts as hollow red diamonds. A = Control, B = Aerobic, C = Strength, D = Aerobic and Strength.

Figure 7. Funnel plot for small-study effects. A = Control, B = Aerobic, C = Strength, D = Aerobic and Strength.

Figure 8.

Cumulative probability curves for the BMI z-score network showing the estimated and predictive probabilities for each treatment being up to a specific rank.

Table 1

Study characteristics.

Study	Year	Country	N	Age (yrs) ($\bar{X} \pm SD$) or Range	Sex
Ackel-D'Elia et al.(59) *	2014	Brazil	AE = 24 AE+ST = 24	AE, AE+ST = 16.5 $\pm$ 1.5	MF
Alberga et al.(60)	2013	Canada	AE+ST = 12 CON = 7	AE+ST = 10.0 $\pm$ 1.0 CON = 10.0 $\pm$ 2.0	MF
Alves et al.(61)	2008	Brazil	AE = 39 CON = 39	AE = 8.0 $\pm$ 1.8 CON = 7.9 $\pm$ 1.5	MF
Cheng et al.(62)	2012	China	AE+ST = 30 CON = 30	AE+ST, CON = 13.0 to 14.0	M
Daley et al.(63)	2006	United Kingdom	AE = 28 CON = 23	AE, CON = 11.0 to 16.0	MF
Davis et al.(64)	2012	United States	AE (LD) = 71 AE (HD) = 73 CON = 78	AE (LD) = 9.3 $\pm$ 0.9 AE (HD) = 9.4 $\pm$ 1.2 CON = 9.4 $\pm$ 1.1	MF
Elloumi et al.(65)	2011	Tunisia	AE = 7 CON = 8	AE = 13.1 $\pm$ 1.0 CON = 13.2 $\pm$ 0.2	M
Farpour-Lambert et al.(66)	2009	Switzerland	AE+ST = 22 CON = 22	AE+ST = 9.1 $\pm$ 1.4 CON = 8.8 $\pm$ 1.6	MF
Fazelifar et al.(67)	2013	Iran	AE+ST = 12 CON = 12	AE+ST, CON = 11.0 to 13.0	M
Ghorbanian et al.(68)	2013	Iran	AE = 15 CON = 15	AE = 17.4 $\pm$ 1.1 CON = 16.9 $\pm$ 1.2	M
Hagstromer et al.(69)	2009	Sweden	AE+ST = 16 CON = 15	AE+ST = 13.7 $\pm$ 2.0 CON = 13.6 $\pm$ 2.2	MF
Karacabey(70)	2009	Turkey	AE = 20 CON = 20	AE = 11.8 $\pm$ 0.5 CON = 11.2 $\pm$ 0.2	M
Kelly et al.(32)	2004	United States	AE = 10 CON = 10	AE = 11.0 $\pm$ 2.0 CON = 11.0 $\pm$ 2.3	MF
Kelly et al.(71)	2015	United States	ST = 13	ST = 15.2 $\pm$ 0.9	MF

Study	Year	Country	N	Age (yrs) ($\bar{X} \pm SD$) or Range	Sex
Kim et al.(72)	2007	Korea, South	CON = 13 AE = 14	CON = 15.5 $\pm$ 0.9 AE = 17 $\pm$ 0.4	M
Kim et al.(73)	2008	Korea, South	CON = 12 AE+ST = 8	CON = 17.0 $\pm$ 0.4 AE+ST, CON = NA	M
Lee et al.(74)	2012	United States	CON = 9 AE = 16 ST = 16	AE = 15.2 $\pm$ 0.9 ST = 14.6 $\pm$ 1.5	M
Lee et al.(75)	2013	United States	CON = 13 AE = 16 ST = 16	CON = 14.8 $\pm$ 1.4 AE = 14.6 $\pm$ 1.9 ST = 14.8 $\pm$ 1.9	F
Li et al.(76)	2014	China	CON = 12 AE = 20 CON = 20	CON = 15.0 $\pm$ 2.2 AE = 15.4 $\pm$ 2.6 CON = 14.6 $\pm$ 3.5	M
Maddison et al.(77)	2011	New Zealand	AE = 160 CON = 162	AE, CON = 11.6 $\pm$ 1.1	MF
Meyer et al.(78)	2006	Germany	AE = 33 CON = 34	AE = 13.7 $\pm$ 2.1 CON = 14.1 $\pm$ 2.4	MF
Murphy et al.(79)	2009	United States	AE = 23 CON = 12	AE = 10.3 $\pm$ 1.9 CON = 10.0 $\pm$ 1.3	MF
Rooney et al.(80)	2005	United States	AE = 26 CON = 33	AE = 8.9 $\pm$ 2.2 CON = 8.6 $\pm$ 2.1	MF
Saygin & Ozturk(81)	2011	Turkey	AE = 20 CON = 19	AE, CON = 10.0 to 12.0	F
Shatibi et al.(82)	2006	United States	ST = 11 CON = 11	ST = 15.1 $\pm$ 1.7 CON = 15.6 $\pm$ 1.7	M
Sigal et al.(83)	2014	Canada	AE = 75 ST = 78 AE+ST = 75	AE = 15.5 $\pm$ 1.4 ST = 15.9 $\pm$ 1.5 AE+ST = 15.5 $\pm$ 1.3	MF
Song et al.(84)	2012	Korea, South	CON = 76 AE = 12 CON = 10	CON = 15.6 $\pm$ 1.3 AE = 12.7 $\pm$ 0.7 CON = 12.6 $\pm$ 0.6	M

Study	Year	Country	N	Age (yrs) ($\bar{X} \pm SD$) or Range	Sex
Sun et al.(85)	2011	China	AE = 25 CON = 17	AE, CON 13.6 $\pm$ 0.7	MF
Tan et al.(86)	2010	China	AE = 30 CON = 30	AE = 9.4 $\pm$ 0.5 CON = 9.5 $\pm$ 0.5	MF
Wagener et al. (87)	2012	United States	AE = 20 CON = 20	AE, CON = 14.0 $\pm$ 1.7	MF
Watts et al.(88)	2004	Australia	AE+ST = 19 CON = 19	AE+ST, CON = 14.3 $\pm$ 1.5	MF
Watts et al.(31)	2004	Australia	AE = 14 CON = 14	AE, CON = 8.9 $\pm$ 1.6	MF
Weintraub et al.(89)	2008	United States	AE = 9 CON = 12	AE = 9.5 $\pm$ 0.6 CON = 10.3 $\pm$ 0.8	MF
Wong et al.(90)	2008	Singapore	AE+ST = 12 CON = 12	AE+ST = 13.8 $\pm$ 1.1 CON = 14.3 $\pm$ 1.5	M

Notes: N, number of participants; yrs, years; AE, aerobic exercise, LPA, leisure physical activity; ST, strength training; CON, control; ($\bar{X} \pm SD$), mean $\pm$ standard deviation; M, males; F, females; NA, not available; M, Males, F, Females; LD, low-dose; HD, high-dose.

* study also in a leisure intervention but was excluded because it didn't meet our eligibility criteria.

Table 2

Initial physical characteristics of participants.

Variable	Exercise				Control			
	# Groups / Participants	$\bar{x} \pm SD$	Mdn	Range	# Groups/Participants	$\bar{x} \pm SD$	Mdn	Range
Age (yrs)	37/1102	13.0 $\pm$ 2.7	14	8 - 17	30/778	12.5 $\pm$ 2.7	13	8 - 17
Height (cm)	27/821	158.0 $\pm$ 12.5	163	130 - 176	23/602	156.4 $\pm$ 12.8	160	127 - 175
Body weight (kg)	33/758	78.7 $\pm$ 18.5	81	35 - 107	27/622	75.8 $\pm$ 18.2	74	34 - 100
BMI z-score	16/730	2.4 $\pm$ 0.6	2.3	1.3 - 2.3	13/508	2.4 $\pm$ 0.6	2.3	1.3 - 3.2
Energy intake (kcal)	8/396	2248 $\pm$ 610	2160	1600-3278	5/181	2355 $\pm$ 754	2134	1614-3259

Notes: Groups (#), number of groups in which data were available; $\bar{x} \pm SD$, mean $\pm$ standard deviation; Mdn, Median; BMI z-score, body mass index z-score; yrs, years; kg, kilograms; kcal, kilocalories.

Table 3

Exercise characteristics.

Variable	# Groups	$\bar{x} \pm SD$	Mdn	Range
Length (weeks)	40	14 ± 6	12	6 - 24
Frequency (days/week)	40	3 ± 1	3	1 - 7
Duration (minutes)	29	47 ± 18	50	6 - 75
Compliance (%)	13	83 ± 22	95	42 - 100
Minutes per week	29	144 ± 53	150	40 - 250
Minutes per week (adjusted)	11	125 ± 45	115	50 - 176
Total minutes	29	2007 ± 1269	1600	720 - 5994
Total minutes (adjusted)	11	1662 ± 633	1915	598 - 2517

Notes: #, number; $\bar{x} \pm SD$, mean $\pm$ standard deviation; Mdn, Median, Compliance (%), percentage of exercise sessions attended; adjusted, adjusted for compliance.

Table 4
Treatment rankings of estimated and predicted probabilities for BMI z-score changes.*

Treatment	SUCRA		PrBest		Mean Rank	
	Est.	Pred.	Est.	Pred.	Est.	Pred.
Aerobic & Strength	76.6	65.9	48.5	39.2	1.7	2.0
Aerobic	73.1	63.0	35.4	33.7	1.8	2.1
Strength	45.3	47.9	16.1	21.6	2.6	2.6
Control	4.9	23.3	0	5.5	3.9	3.3

Notes:

* Treatments listed from best to worst; Est., estimated; Pred., predicted; SUCRA, surface under the cumulative ranking analysis; PrBest, Probability of being the best treatment.

Table 5

Summary of confidence in effect estimates and ranking of treatments for BMI z-score.

Comparison	Evidence Type	Confidence	Reasons for Downgrading
Aerobic vs. Control	Mixed	Low	Study limitations ^a ; Inconsistency ^b
Strength vs. Control	Mixed	Very low	Study limitations ^a ; Inconsistency ^b ; Imprecision ^c
Aerobic & Strength vs. Control	Mixed	Very low	Study limitations ^a ; Inconsistency ^b ; Publication bias ^d
Strength vs Aerobic	Mixed	Very low	Study limitations ^a ; Inconsistency ^b ; Imprecision ^c
Aerobic & Strength vs. Aerobic	Mixed	Very low	Study limitations ^a ; Inconsistency ^b ; Imprecision ^c
Aerobic & Strength vs Strength	Mixed	Very low	Study limitations ^a ; Inconsistency ^b ; Imprecision ^c
Ranking of treatments	NA	Low	Study limitations ^a ; Inconsistency ^b

Notes: NA, not applicable; Confidence rankings = high, moderate, low, very low;

^aDominated by evidence at high or unclear risk of bias;

^bOverlapping 95% prediction intervals;

^cOverlapping 95% confidence intervals;

^dFunnel plot asymmetry; Number of observations for each comparison were 24 (Aerobic versus Control), 5 (Strength versus Control), 9 (Aerobic and Strength versus Control), 3 (Strength versus Aerobic), 2 (Aerobic and Strength versus Aerobic) and 1 (Aerobic and Strength versus Strength).