Review Article

Exercise and BMI in Overweight and Obese Children and Adolescents: A Systematic Review and Trial Sequential Meta-Analysis

George A. Kelley, Kristi S. Kelley, and Russell R. Pate

1Department of Biostatistics, School of Public Health, West Virginia University, Morgantown, WV 26506, USA
2Department of Exercise Science, University of South Carolina, Columbia, SC 29208, USA

Correspondence should be addressed to George A. Kelley; gkelley@hsc.wvu.edu

Received 13 June 2015; Revised 11 September 2015; Accepted 13 September 2015

Academic Editor: Giedrius Vanagas

Copyright © 2015 George A. Kelley et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Objective. Determine the effects of exercise on body mass index (BMI in kg/m^2) among overweight and obese children and adolescents. Methods. Trial sequential meta-analysis of randomized controlled exercise intervention trials ≥ 4 weeks and published up to November 11, 2014. Results. Of the 5,436 citations screened, 20 studies representing 971 boys and girls were included. Average length, frequency, and duration of training were 13 weeks, 3 times per week, for 46 minutes per session. Overall, random-effects models showed that exercise decreased BMI by 3.6% (mean: -1.08; 95% CI: -0.52 to -1.64; Q = 231.4; p < 0.001; I^2 = 90.9%; 95% CI: 87.6% to 93.4%; D^2 = 91.5%). Trial sequential meta-analysis showed that changes in BMI crossed the monitoring boundary for a type I error in 2010, remaining stable thereafter. The number needed to treat was 5 while the percentile improvement was 26.9. It was estimated that approximately 2.5 million overweight and obese children in the US and 22.0 million overweight and obese children worldwide could reduce their BMI by participating in a regular exercise program. Overall quality of evidence was rated as moderate. Conclusions. Exercise is associated with improvements in BMI among overweight and obese children and adolescents. This trial is registered with PROSPERO Trial Registration #CRD42015017586.

1. Introduction

The prevalence of overweight and obesity in children and adolescents is a pandemic problem both in the United States (US) and worldwide. Recently, Ogden et al. reported that the prevalence of overweight and obesity in the US, defined as a body mass index (BMI) in kg/m^2 ≥ 85th percentile based on Centers for Disease Control Growth Charts, was 31.8% among children and adolescents 2 to 19 years of age, while the prevalence of obesity, defined as a BMI in kg/m^2 ≥ 95th percentile, was 16.9% [1]. When compared to 30 years ago, this represents an obesity prevalence that is more than two times higher in US children and more than four times higher in adolescents [1, 2]. From a worldwide perspective, the prevalence of overweight and obesity in 2013 has been reported to be approximately 23% among children and adolescents in developed countries and 13% among children and adolescents from developing countries [3]. Collectively, this represents an approximate 47% increase in the worldwide prevalence of overweight and obesity among children and adolescents between 1980 and 2013 [3].

The economic costs associated with overweight and obesity among children and adolescents are also substantial. For example, Finkelstein et al. estimated that the incremental lifetime medical cost of an obese 10-year-old child in the US, in relation to a normal weight child who maintained normal weight throughout adulthood, was $19,000 [4]. Based on the current number of obese 10-year-olds in the US, the total direct medical costs associated with obesity were estimated at $14 billion for this age only [4].

The negative health consequences of obesity in children and adolescents are both immediate and long-term. For example, in a population-based sample of US children and adolescents 5 to 17 years of age from the Bogalusa Heart Study,
approximately 70% of obese youth had at least one cardiovascular disease risk factor [5]. In addition, obese children and adolescents, in relation to their normal weight peers, suffer from a greater prevalence and/or incidence of other conditions that include, but are not necessarily limited to, musculoskeletal pain, injuries and fractures [6], obstructive sleep apnea [7], and poorer self-esteem and quality of life [8]. From a long-term perspective, overweight and obesity during childhood and adolescence have been shown to track into adulthood [9], thereby placing this population group at an increased risk for premature all-cause mortality [10]. This is a major problem since overweight and obesity have been reported to be the third leading cause of preventable death in the US, responsible for 216,000 deaths in 2005 [10]. Globally, the World Health Organization has estimated that approximately 3.4 million adults die each year as a result of being overweight or obese [11]. The issue of obesity has become so problematic that it is now recognized by the American Medical Association as a disease [12].

Exercise has been recommended for the prevention and treatment of overweight and obesity in children and adolescents [13–18]. In a recent systematic review with meta-analysis of studies published until the year 2012, the investigative team reported a statistically significant decrease of approximately 3% in BMI z-score in overweight and obese children and adolescents [19]. However, body mass index BMI in kg m$^{-2}$ continues to be the most commonly assessed and reported metric and is easily recognized and interpreted by practitioners. Unfortunately, the effects of exercise on BMI in kg m$^{-2}$ have been underwhelming. For example, with the exception of one previous systematic review with meta-analysis that focused on exercise [20], others reported a nonsignificant decrease in BMI in kg m$^{-2}$ among children and adolescents [17, 21–23]. However, all five suffer from potential limitations. These include (1) the pooling of a small number of exercise-only studies [17, 21, 22], (2) the inclusion of nonrandomized trials [20, 22], (3) inclusion of children and adolescents who were not overweight or obese [20, 22, 23], and (4) overall quality scores ranging from only 45% to 82% when the Assessment of Multiple Systematic Reviews (AMSTAR) instrument was applied to the studies [24]. In addition, none of the studies used trial sequential analysis, an approach that can provide data regarding (1) adequate information size, (2) a threshold for a statistically significant effect, and (3) a threshold for futility [25]. Given the former, the purpose of the current study was to conduct a systematic review and trial sequential meta-analysis of randomized controlled trials addressing the overall effects of exercise (aerobic training, strength training, or both) on BMI in kg m$^{-2}$ among overweight and obese children and adolescents.

2. Methods

2.1. Registration and General Procedure. This systematic review with trial sequential meta-analysis is registered in PROSPERO (#CRD42015017586), an international prospective registry of systematic reviews. The conduct and reporting of this study was accomplished according to the general guidelines recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [26].

2.2. Study Eligibility. The a priori inclusion criteria for this study were as follows: (1) randomized controlled trials (assignment at participant level only), (2) control group (nonintervention, usual care, wait-list control, and attention control), (3) exercise (aerobic training, strength training, or both) ≥ 4 weeks as an independent intervention, (4) overweight and obese children and adolescents, as defined by the authors, (5) boys and/or girls 2 to 18 years of age, (6) studies published in full in any language between January 1, 1990, and November 11, 2014, and (7) data available for calculating changes in BMI in kg m$^{-2}$. Studies were excluded based on an inappropriate population, intervention, comparison, outcome, study type, or lack of requisite data for BMI in kg m$^{-2}$.

2.3. Data Sources. The following databases were searched from January 1, 1990, to December 31, 2012: (1) Academic Search Complete, (2) CINAHL, (3) Cochrane Central Register of Controlled Trials (CENTRAL), (4) Education Research Complete, (5) ERIC, (6) LILACS, (7) Medline, (8) Proquest, (9) Scopus, (10) Sport Discus, and (11) Web of Science. In addition, an updated PubMed search was conducted for potentially eligible studies published between August 1, 2012, and November 11, 2014. A brief description of each database is shown in Supplementary File 1 (see Supplementary Material available online at http://dx.doi.org/10.1155/2015/704393) while the updated search strategy for PubMed can be found in Supplementary File 2. Database searches were supplemented by cross-referencing for potentially eligible studies, including reviews, as well as expert review by the third author. All studies were stored in Reference Manager, version 12.0 [27]. Overall precision of the searches was computed by dividing the number of studies included by the total number of studies screened while the number needed to read (NNR) was calculated as the inverse of the precision [28].

2.4. Study Selection. Independent, dual-selection of eligible studies was conducted by the first two authors who then met and reviewed their choice for inclusion. Disagreements were resolved by consensus and, if necessary, consultation with the third author.

2.5. Data Abstraction. Codebooks were developed in an electronic spreadsheet program [29] that included items that fell within the following four major categories: (1) study characteristics, (2) physical characteristics of participants, (3) training program characteristics, and (4) outcomes and outcome characteristics. Independent, dual-selection of eligible studies was conducted by the first two authors who then met and reviewed their choice for inclusion. Disagreements were resolved by consensus and, if necessary, consultation with the third author. Using Cohen’s kappa statistics (κ) [30], the
overall agreement rate prior to correcting discrepancies was 0.94.

2.6. Risk of Bias Assessment. The Cochrane Risk of Bias Assessment Instrument was used to assess potential risk of bias [31]. Items were rated as being at a high, low, or unclear risk for bias [31]. In addition to the six basic items, an additional item about whether the participants were exercising regularly prior to study participation, as defined by the original study authors, was included [31]. Assessment for risk of bias was limited to the primary outcome, BMI in kg⁻². Dual and independent assessment for risk of bias was conducted by the first two authors who subsequently met and reviewed every item for agreement. Disagreements were resolved by consensus and, if necessary, consultation with the third author. Using Cohen’s kappa statistics (κ) [30], the overall agreement rate prior to correcting discrepancies was 0.70.

2.7. Data Synthesis

2.7.1. Calculation of Effect Sizes for BMI in kg⁻². The primary outcome for this study was changes in BMI in kg⁻². Secondary outcomes included body weight, percent body fat, fat mass, fat-free mass, changes in maximum oxygen consumption in ml·kg⁻¹·min⁻¹ (VO₂max), and upper and lower body strength and kilocalorie intake. Effect sizes (ES) using the original metrics were calculated by subtracting the change score difference in the exercise group from the change score 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 reported, they were calculated from pre- and poststandard deviations according to procedures developed by Follmann et al. [32]. Each ES was then weighted by the inverse of its variance.

2.7.2. Pooled Estimates for Changes in BMI in kg⁻². Changes in BMI in kg⁻² and all secondary outcomes were pooled using random-effects, method-of-moments models that incorporate between-study heterogeneity into the final estimate [33]. Ninety-five percent confidence intervals (CI) were calculated while z-based two-tailed alpha values ≤ 0.05 were considered statistically significant. Heterogeneity was examined using the Q statistic [34], with an alpha value ≤ 0.10 representative of statistically significant heterogeneity. Inconsistency was examined using I² [35] and diversity using D² [36]. For both I² and D² values < 25%, 25% to <50%, 50% to <75%, and 75% or greater were considered to represent very low, low, moderate, and large amounts of inconsistency and diversity [37]. Statistically significant outliers were considered to be those with standardized residual alpha values ≤ 0.05. Multiple exercise groups in the same study were analyzed independently as well as collapsing multiple groups so that only one ES represented each study while the sample size for the control group was divided by the number of exercise groups [38]. In addition to 95% CI, 95% prediction intervals (PI) were also calculated [39, 40] for any result that was statistically significant. Based on recent recommendations [41], small-study effects (publication bias, etc.) were examined both qualitatively and quantitatively using funnel plots and Egger’s regression intercept test [42]. A one-tailed probability value < 0.05 was considered to be indicative of statistically significant small-study effects. The influence of each result on the overall findings was examined by deleting each result from the model once. Post hoc, the fail-safe N test was used to estimate the number of studies that would be needed to reverse our finding of a statistically significant, that is, p < 0.05, improvement in BMI in kg⁻² [43]. This test was used because four studies that met all of our inclusion criteria except for the provision of sufficient postintervention data were excluded from the meta-analysis.

To enhance practical application, the number needed to treat (NNT) was calculated for changes in BMI in kg⁻² assuming a conservative control group risk of 10% and only if changes in BMI in kg⁻² were statistically significant. If the NNT was calculated, gross estimates were determined for the number of obese children and adolescents in the US as well as worldwide that could potentially benefit from exercise. These estimates were based on 12.5 [1] and 110 million [44, 45] overweight and obese children in the US and worldwide, respectively. In addition to NNT, Cohen’s U₃ index, an index used to determine the percentile gain in an intervention group, was calculated for any statistically significant results with respect to BMI in kg⁻² and secondary outcomes [46]. Finally, The Grades of Recommendation, Assessment, Development and Evaluation (GRADE) instrument was used to assess the overall quality of evidence and was limited to the primary outcome, BMI in kg⁻² [47]. Overall quality was categorized as very low, low, moderate, or high [47].

Based on empirical evidence that consideration of information size and adjusted significance thresholds may avoid false statistical inferences due to imprecision and repeated significance testing in meta-analysis [25, 48–50], information size estimates and trial sequential analysis were performed [51] for BMI in kg⁻². Trial sequential analysis is an approach that combines conventional meta-analysis methodology with meta-analytic sample size considerations as well as previously established methods for repeated significance testing on accruing data in randomized trials [51]. Inferences derived from using trial sequential analysis may be more reliable than using conventional meta-analysis procedures [51]. More specifically, previous research suggests that information size considerations as well as adjusted significance thresholds may eliminate early false positive findings due to a lack of precision and repeated significance testing in meta-analyses [25, 48–51].

The a priori plan was to estimate the required information size based on previous research suggesting that a 0.1 kg/m² change in BMI in kg⁻² can be clinically important [52]. However, because of the inability to obtain variance statistics, a post hoc decision was made to estimate the required information size using the pooled mean difference and variance, adjusted for between-study heterogeneity, from the current study. A two-tailed type I error rate of 5% and power of 80% were employed. To control for multiple tests, trial sequential...
monitoring boundaries for both type 1 (5%) and type 2 (20%) error rates were established using O'Brien-Fleming adjustments [53, 54].

2.7.3. Metaregression Analysis. Simple, random-effects metaregression (method of moments) models were used to examine associations between changes in BMI in kg·m⁻² and potential predictors [33]. An a priori decision was made to not conduct any type of multiple metaregression analyses because of missing data for different variables from different studies. Metaregression analysis was limited to those studies in which there were at least four results for continuous variables or four results per group for categorical variables. Continuous variables, determined a priori, included year of publication, percent dropout, age, baseline BMI in kg·m⁻², and exercise intervention (length, frequency, duration, compliance, minutes per week, unadjusted and adjusted for compliance, and total minutes for the intervention, unadjusted and adjusted for compliance). Categorical variables examined included country, type of control group, funding, a priori sample size estimates, adverse events, risk of bias (sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and whether subjects were inactive prior to enrollment), gender, race/ethnicity, changes in exercise and/or physical activity outside the exercise intervention, pubertal stage, type of exercise (aerobic, strength, and both), exercise supervision, setting that exercise took place, type of participation, type of analysis, and exercise intensity (low, moderate, and high) [55].

2.7.4. Reporting and Software Utilization. Changes in primary and secondary outcomes are reported in their natural direction of benefit, that is, negative values for changes in BMI in kg·m⁻² and positive values for increases in fat-free mass. All statistical analyses were conducted using Comprehensive Meta-Analysis (version 3.3) [56], Microsoft Excel 2010 [57], Trial Sequential Analysis (version 0.9) [51], GradePro (version 3.6) [58], and two add-ins for Microsoft Excel, SSC-stat (version 2.18) [59] and EZ-Analyze (version 3.0) [60].

3. Results

3.1. Characteristics of Included Studies. After removing duplicates, a total of 5,436 articles were screened. Of these, 20 studies representing 42 groups (22 exercise, 20 control) and final assessment of BMI in kg·m⁻² in 971 participants (575 exercise, 396 control) met all eligibility criteria [61–80]. Overall precision of the searches was 0.004 while the NNR was 272. The major reasons for exclusion were inappropriate study design (51.6%), intervention (31.2%), population (14.0%), comparison (2.3%), and outcome(s) (0.9%). Another four studies comprising less than 1% of the reasons for exclusion were omitted because data necessary for conducting trial sequential meta-analysis were not available [81–84]. This included (1) lack of both post and change outcome values as well as standard deviations, or data for conversion to standard deviations (e.g., standard error of the mean), for BMI in kg·m⁻² [81–83] and (2) lack of separate sample sizes for exercise and control groups [84]. A flow diagram of the search process is shown in Figure 1 while a list of excluded studies, including the specific reason(s) for exclusion, can be found in Supplementary File 3.

Table 1 describes the characteristics of each included study. Studies were conducted in 12 different countries and published between 2004 and 2014 [61–80], with all but one [62] published in English-language journals. Assessment of primary and secondary outcomes took place after six [68, 69], eight [63, 67, 77–79], 10 [76], 12 [61, 64, 66, 71, 72, 75, 80], 13 [65], 16 [73, 74], and 24 [62, 70] weeks of exercise. Two studies used some type of matching procedure, one according to age, gender, and BMI in kg·m⁻² [67] and another according to sex and degree of overweight [74]. For those studies in which it could be determined, six used the per-protocol approach to analyze their data [61, 65, 69, 70, 73, 75], two used intention-to-treat [76, 77], and three used both [62, 64, 74]. Only five of the studies included sample size estimates [62, 64, 74, 75, 79] while the majority (80%) received some type of funding for their work [61, 63–65, 67–71, 73–76, 78–80]. Nine studies received singular support from either university [69, 70, 75, 80], government [63, 65, 76], or private [78, 79] entities while seven other studies reported multiple sources of support from government and private [67, 71], government and university [64, 68], government, university, and private [73, 74], or private and university [61] entities.

For those studies in which data were available, the dropout rate for studies in which data were available ranged from 0% to 34% in the exercise groups (X ± SD, 16.9% ± 14.0, median = 23) and 0% to 26% in the control groups (X ± SD, 12.6% ± 12.0, median = 14). Reasons for dropping out included time, lack of interest, unhappiness with group assignment, moving, and medical condition. Four studies reported no serious adverse events during the intervention period [74, 77–79].

Participant characteristics are shown in Tables 1 and 2. Twelve of the 20 studies (60%) included both boys and girls [61, 62, 64, 65, 67, 70, 71, 74, 76–79], seven (35%) were limited to boys [63, 66, 68, 69, 73, 75, 80], and one (5%) was limited to girls [72]. Participants included Whites, Blacks, Asians, and Hispanics. With respect to maturational development, the studies represented boys and girls at the prepubertal, pubertal, and postpubertal stages of development; two studies reported boys and girls at the prepubertal stage [61, 64], one at the postpubertal stage [74], two at the prepubertal and pubertal stage [67, 76], and one each at either the pubertal and postpubertal [73] or prepubertal, pubertal, and postpubertal [63] stages. Characteristics of the exercise programs from each study are also shown in Table 1. Thirteen of the 22 groups participated in aerobic exercise, two in strength training, and seven in both. For those studies and groups in which data were available, length of training ranged from six to 24 weeks (X ± SD, 13.4 ± 5.7, median = 12) and frequency from one to five times per week (X ± SD, 3.3 ± 1.1, median = 3). Intensity of training was classified as low for one group, moderate for
four groups, moderate to high for two groups, and high for seven groups. When limited to aerobic exercise, duration of training for the 18 groups in which data were available ranged from 20 to 75 minutes per session ($\bar{X} \pm SD, 45.9 \pm 15.4$, median = 45). Total minutes of training, per week, ranged from 40 to 224 minutes ($\bar{X} \pm SD, 148.1 \pm 55.0$, median = 155) while total minutes of training for the entire intervention period ranged from 480 to 5,040 minutes ($\bar{X} \pm SD, 1979 \pm 1302$, median = 1540). When adjusted for compliance to the exercise protocol, total minutes per week for the three groups in which data could be calculated ranged from 39 to 75 minutes ($\bar{X} \pm SD, 67.3 \pm 25.2$, median = 75) while total minutes over the entire intervention period ranged from 470 to 896 minutes ($\bar{X} \pm SD, 749 \pm 241$, median = 880). Aerobic exercises included walking, jogging, cycling, swimming, jumping rope, stair climbing, aerobic dance, and games (soccer, handball, basketball, volleyball, etc.) as well as other various activities.

For strength training groups, the within-study number of sets for the five groups in which data were provided ranged from one to three while the number of repetitions per set for the six groups in which data were available ranged from three to 25. Two strength training groups reported within-study rest periods between sets that ranged between 60 and 180 seconds. For the five groups that reported data, the number of strength training exercises ranged from seven to 13 ($\bar{X} \pm SD, 9.2 \pm 2.5$, median = 9). Types of resistance training equipment used included free weights, machine weights, elastic bands, medicine balls, and the participants' own body weight. Across all exercise groups, 18 participated in supervised exercise, one in unsupervised exercise, and three in both. Compliance to the exercise interventions for the four groups in which data could be calculated ranged from 55% to 98% ($\bar{X} \pm SD, 83.4 \pm 20.2$, median = 90).

3.2 Risk of Bias Assessment. The results for pooled risk of bias assessment are shown in Figure 2 while study level results are shown in Supplementary File 4. As can be seen, 95% of the included studies adequately described the process for random sequence generation while none of the studies suffered from incomplete outcome reporting. In contrast, more than half of
<table>
<thead>
<tr>
<th>Study</th>
<th>Year</th>
<th>Sex</th>
<th>N</th>
<th>Age</th>
<th>Length</th>
<th>Frequency</th>
<th>Intensity (%)</th>
<th>Duration (min)</th>
<th>Mode</th>
<th>Compliance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alberga et al. [61]</td>
<td>2013</td>
<td>M/F</td>
<td>12/7</td>
<td>10.0 ± 1.0/10.0 ± 2.0</td>
<td>12</td>
<td>2</td>
<td>65–70</td>
<td>65–86</td>
<td>20</td>
<td>45</td>
</tr>
<tr>
<td>Alves et al. [62]</td>
<td>2008</td>
<td>M/F</td>
<td>39/39</td>
<td>8.0 ± 1.8/7.9 ± 1.5</td>
<td>24</td>
<td>3</td>
<td>—</td>
<td>—</td>
<td>50</td>
<td>—</td>
</tr>
<tr>
<td>Elloumi et al. [63]</td>
<td>2011</td>
<td>M</td>
<td>7/8</td>
<td>13.1 ± 1/13.2 ± 0.2</td>
<td>8</td>
<td>4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Farpour-Lambert et al. [64]</td>
<td>2009</td>
<td>M/F</td>
<td>22/22</td>
<td>9.1 ± 1.4/8.8 ± 1.6</td>
<td>12</td>
<td>3</td>
<td>—</td>
<td>55–65</td>
<td>30</td>
<td>20</td>
</tr>
<tr>
<td>Hagstrømer et al. [65]</td>
<td>2009</td>
<td>M/F</td>
<td>16/15</td>
<td>13.7 ± 2.0/13.6 ± 2.2</td>
<td>13</td>
<td>1</td>
<td>—</td>
<td>50–70</td>
<td>60</td>
<td>—</td>
</tr>
<tr>
<td>Karacabey [66]</td>
<td>2009</td>
<td>M</td>
<td>20/20</td>
<td>11.8 ± 0.5/11.2 ± 0.80</td>
<td>12</td>
<td>3</td>
<td>60–65</td>
<td>—</td>
<td>20–45</td>
<td>—</td>
</tr>
</tbody>
</table>
| Kelly et al. [67]     | 2004 | M/F | 10/10 | 11.0 ± 0.6/11.0 ± 0.7 | 8  | 4 | — | 50–80 | 30–50 | — | A | ≥80%
| Kim et al. [68]       | 2007 | M   | 14/12 | 170 ± 0.4/170 ± 0.4 | 6  | 5 | — | — | 30 | — | A | — |
| Kim et al. [69]       | 2008 | M   | 8/9  | 11.0/11.0 | 12 | 2 | 55–75 | 70 | 30–35 | 50 | A + S | — |
| Meyer et al. [70]     | 2006 | M/F | 33/34 | 13.7 ± 2.1/14.1 ± 2.4 | 24 | 3 | — | — | 60–90 | — | A | — |
| Murphy et al. [71]    | 2009 | M/F | 23/12 | 7–12 | 12 | 5 | — | — | 10–30 | — | A | 75% ≥ 5x weeks; 15% ≥ 3x weeks |
| Saygin and Öztürk [72]| 2011 | F   | 20/19 | 10–12 | 12 | 3 | 43 | — | 75 | — | A | — |
| Shaibi et al. [73]    | 2006 | M   | 11/11 | 15.1 ± 0.5/15.6 ± 0.5 | 16 | 2 | — | 62–97 | — | — | S | 96% |
| Sigal et al. [74]     | 2014 | M/F | 75 (A) | 15.5 ± 1.4 (A) | 22 | 4 | 65–85 | — | 20–50 | — | A | 62% (A) |
|                       |      |     | 78 (S) | 15.9 ± 1.5 (S) | — | — | 65–85 | — | 20–50 | — | A | 56% (S) |
|                       |      |     | 75 (A + S) | 15.5 ± 1.3 (A + S) | — | — | 65–85 | — | 20–50 | A + S | 64% (A + S) |
|                       |      |     | 76 (Con) | 15.6 ± 1.3 (Con) | — | — | — | — | 55 | — | — |
| Song et al. [75]      | 2012 | M   | 12/10 | 12.7 ± 0.7/12.6 ± 0.6 | 12 | 3 | 60–70 | — | — | 30 | — | A | ≥80% |
| Sun et al. [76]       | 2011 | M/F | 25/17 | 13.6 ± 0.7 | 10 | 4 | — | 40–60 | 40 | — | — | 55% |
| Tan et al. [77]       | 2010 | M/F | 30/30 | 9.4 ± 0.5/9.5 ± 0.5 | 8  | 5 | — | — | — | — | A | — |
| Watts et al. [78]     | 2004 | M/F | 19/19 | 14.3 ± 1.5 | 8  | 3 | 65–85 | — | 55–70 | 60 | A + S | ≥90% |
| Watts et al. [79]     | 2004 | M/F | 14/14 | 8.9 ± 1.6 | 8  | 3 | 66–85 | — | — | 60 | — | A | ≥90% |
| Wong et al. [80]      | 2008 | M   | 12/12 | 13.8 ± 1/14.3 ± 1.5 | 12 | 2 | 50–85 | — | — | 55 | — | A + S | — |

Notes. M/F: males/females; Ex/Con: exercise/control; days: days per week; minutes: minutes per session; A/S: aerobic/strength; compliance: percentage of exercise sessions attended; data reported as mean ± standard deviation; VO₂max: maximum oxygen consumption in mL·kg⁻¹·min⁻¹; IRM: one-repetition maximum; MHRR: maximal heart rate reserve; MHR: maximum heart rate; *compliance based on medians versus means; b participants trained at intensity equivalent to 78% of lactate threshold; c study consisted of circuit training.
### Table 2: Initial physical characteristics of participants.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Exercise Groups/participants</th>
<th>Missing data (%)</th>
<th>Mean ± SD</th>
<th>Mdn</th>
<th>Range</th>
<th>Control Groups/participants</th>
<th>Missing data (%)</th>
<th>Mean ± SD</th>
<th>Mdn</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>21/489</td>
<td>4.5</td>
<td>12.4 ± 2.5</td>
<td>13</td>
<td>8–17</td>
<td>18/301</td>
<td>10.0</td>
<td>12.0 ± 2.6</td>
<td>12</td>
<td>8–17</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>19/443</td>
<td>13.6</td>
<td>155.7 ± 12.5</td>
<td>163</td>
<td>130–174</td>
<td>17/263</td>
<td>15.0</td>
<td>154.7 ± 13.4</td>
<td>154</td>
<td>127–175</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>20/370</td>
<td>9.1</td>
<td>74.5 ± 18.7</td>
<td>75</td>
<td>35–104</td>
<td>18/312</td>
<td>10.0</td>
<td>73.3 ± 18.5</td>
<td>74</td>
<td>34–99</td>
</tr>
<tr>
<td>BMI (kg m⁻²)</td>
<td>22/433</td>
<td>0.0</td>
<td>29.7 ± 4.0</td>
<td>30</td>
<td>21–36</td>
<td>20/376</td>
<td>0</td>
<td>29.6 ± 3.9</td>
<td>30</td>
<td>21–36</td>
</tr>
<tr>
<td>Fat mass (kg)</td>
<td>14/228</td>
<td>36.4</td>
<td>33.4 ± 9.5</td>
<td>32</td>
<td>22–50</td>
<td>12/181</td>
<td>40.0</td>
<td>31.8 ± 7.1</td>
<td>31</td>
<td>22–47</td>
</tr>
<tr>
<td>Body fat (%)</td>
<td>17/321</td>
<td>22.7</td>
<td>39.0 ± 6.2</td>
<td>36</td>
<td>31–50</td>
<td>15/265</td>
<td>25.0</td>
<td>38.5 ± 5.0</td>
<td>38</td>
<td>31–48</td>
</tr>
<tr>
<td>Fat-free mass (kg)</td>
<td>15/261</td>
<td>31.8</td>
<td>45.2 ± 7.9</td>
<td>47</td>
<td>28–54</td>
<td>13/208</td>
<td>35.0</td>
<td>44.9 ± 9.4</td>
<td>46</td>
<td>26–62</td>
</tr>
<tr>
<td>VO₂max (mL·kg⁻¹·min⁻¹)</td>
<td>12/229</td>
<td>40.0</td>
<td>30.2 ± 6.1</td>
<td>31</td>
<td>20–39</td>
<td>10/175</td>
<td>50.0</td>
<td>30.0 ± 6.8</td>
<td>29</td>
<td>20–40</td>
</tr>
<tr>
<td>Muscular strength (kg)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>3</td>
<td>66.7</td>
<td>86.8 ± 84.1</td>
<td>39</td>
<td>37–18</td>
<td>2</td>
<td>75.0</td>
<td>124.4 ± 36.6</td>
<td>124</td>
<td>37–212</td>
</tr>
<tr>
<td>Lower</td>
<td>3</td>
<td>66.7</td>
<td>84.3 ± 35.7</td>
<td>100</td>
<td>43–110</td>
<td>2</td>
<td>75.0</td>
<td>80.0 ± 34.2</td>
<td>80</td>
<td>56–104</td>
</tr>
<tr>
<td>Energy intake (kcals)</td>
<td>6/108</td>
<td>72.7</td>
<td>2465 ± 577</td>
<td>2319</td>
<td>1813–3278</td>
<td>4/44</td>
<td>81.8</td>
<td>2511 ± 771</td>
<td>2586</td>
<td>1614–3259</td>
</tr>
</tbody>
</table>

*Notes. Groups/participants: number of groups and participants in which data were available; Mean ± SD: mean ± standard deviation; Mdn: median; BMI: body mass index; VO₂max: maximum oxygen consumption; kcals: kilocalories; data limited to those groups in which aerobic exercise was an intervention; data limited to those groups in which strength training was an intervention; percentage of missing data calculated based on the premise that (1) all studies should have assessed and reported data for age, height, body weight, BMI, fat mass, percent body fat, fat-free mass, and energy intake, (2) all studies that included an aerobic exercise component should have assessed and reported data on VO₂max in mL·kg⁻¹·min⁻¹, and (3) all studies that included a strength training component should have assessed and reported data on upper and lower body strength.*
the studies were at a high or unclear risk of bias with respect to allocation concealment (85%), blinding of participants and personnel (100%), blinding of outcome assessors (90%), incomplete outcome data, that is, attrition bias (70%), and boys and girls being physically inactive prior to enrollment (70%).

3.3 Data Synthesis

3.3.1 Primary Outcome. Pooled results for changes in BMI in kg·m⁻² are shown in Table 3 and Figure 3. Across all categories, a statistically significant reduction equivalent to 3.6% was found for BMI in kg·m⁻² along with statistically significant heterogeneity, a large amount of inconsistency and diversity, and overlapping prediction intervals. Changes in BMI in kg·m⁻² ranged from 0.59 to −7.30 kg/m². With one outlier deleted from the model [66], reductions were not as large (26% difference) but remained statistically significant along with statistically significant heterogeneity as well as a large amount of inconsistency and diversity when results were collapsed so that only one result represented each

### Table 3: Changes in primary and secondary outcomes.

<table>
<thead>
<tr>
<th>Variable</th>
<th>ES (#)</th>
<th>X (95% CI)</th>
<th>Z (p)</th>
<th>Q (p)</th>
<th>I² (%) (95% CI)</th>
<th>D² (%)</th>
<th>95% PI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI (kg·m⁻²)</td>
<td>22</td>
<td>−1.08 (−0.52,−1.64)*</td>
<td>−3.81 (&lt;0.001)</td>
<td>231.4 (&lt;0.001)*</td>
<td>90.9 (87.6, 93.4)</td>
<td>91.5</td>
<td>−3.74, 1.58</td>
</tr>
<tr>
<td>Secondary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>20</td>
<td>−1.66 (−0.87,−2.45)*</td>
<td>−4.11 (&lt;0.001)</td>
<td>57.3 (&lt;0.001)*</td>
<td>66.8 (47.0, 79.2)</td>
<td>78.0</td>
<td>−4.48, 1.17</td>
</tr>
<tr>
<td>Fat mass (kg)</td>
<td>14</td>
<td>−1.07 (−0.36,−1.79)*</td>
<td>−2.93 (0.003)</td>
<td>29.5 (0.006)*</td>
<td>55.9 (19.8, 75.8)</td>
<td>62.6</td>
<td>−3.33, 1.19</td>
</tr>
<tr>
<td>Body fat (%)</td>
<td>17</td>
<td>−1.13 (−0.58,1.67)*</td>
<td>−4.05 (&lt;0.001)</td>
<td>56.2 (&lt;0.001)*</td>
<td>71.5 (53.6, 82.6)</td>
<td>77.1</td>
<td>−3.16, 0.91</td>
</tr>
<tr>
<td>Fat-free mass (kg)</td>
<td>15</td>
<td>−0.006 (−0.24,0.22)</td>
<td>−0.05 (0.96)</td>
<td>21.8 (0.08)</td>
<td>35.8 (0.65, 4.3)</td>
<td>39.2</td>
<td>—</td>
</tr>
<tr>
<td>VO₂max (mL·kg⁻¹·min⁻¹)³</td>
<td>10</td>
<td>3.1 (1.1, 5.2)*</td>
<td>2.95 (&lt;0.001)</td>
<td>197.8 (&lt;0.001)*</td>
<td>95.4 (93.3, 96.9)</td>
<td>96.8</td>
<td>−4.53, 10.76</td>
</tr>
<tr>
<td>Muscular strength (kg)⁶</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>3</td>
<td>7.7 (4.4, 10.9)*</td>
<td>4.58 (&lt;0.001)</td>
<td>0.9 (0.63)</td>
<td>0 (0, 92.7)</td>
<td>0</td>
<td>6.3, 9.0*</td>
</tr>
<tr>
<td>Lower</td>
<td>3</td>
<td>44.2 (29.5, 59.0)*</td>
<td>5.88 (&lt;0.001)</td>
<td>0.3 (0.86)</td>
<td>0 (0, 77.3)</td>
<td>0</td>
<td>42.9, 45.6*</td>
</tr>
<tr>
<td>Energy intake (kcs)</td>
<td>6</td>
<td>−141 (−294, 13)</td>
<td>−1.80 (0.07)</td>
<td>8.9 (0.11)</td>
<td>43.6 (0, 77.7)</td>
<td>59.4</td>
<td>—</td>
</tr>
</tbody>
</table>

Notes:*Number; ES: effect size; X (95% CI): mean and 95% confidence interval; Z(p): Z value and alpha value for Z; Q(p): Cochrane’s Q statistic and alpha value for Q; I² (%) : I-squared; 95% PI: 95% prediction intervals; D²: D-squared; BMI: body mass index; VO₂max: maximum oxygen consumption; kcs: kilocalories; *data limited to those groups in which aerobic exercise was an intervention; †data limited to those groups in which strength training was an intervention; * statistically significant; —: not calculated; boldface items indicate statistical significance.
study (\(X\); \(-1.10\); 95% CI: \(-0.52\) to \(-1.68\); \(z = -3.71\); \(p < 0.001\); 
\(Q = 230.3\); \(p < 0.001\); \(I^2 = 91.8\%\); 95% CI = 88.7 to 
94.0; \(D^2 = 92.2\%\)). No small-study effects were observed as 
indicated by a lack of funnel plot asymmetry (Figure 4) and 
Egger’s regression-intercept test (\(\hat{\beta}_0 = -0.92\); \(p = 0.34\)). With 
each result deleted from the model once, reductions in BMI in 
kg\(\cdot\)m\(^{-2}\) remained statistically significant across all deletions, 
with changes ranging from \(-0.80\) to \(-1.16\), a difference of 
31% (Figure 5). The NNT was 5 (95% CI = 3 to 12) while the 
percentile improvement was 26.9 (95% CI = 15.3 to 36.0). It 
was estimated that approximately 2.5 million overweight 
and obese children in the US (95% CI, 1.0 to 4.2) and 22.0 

million overweight and obese children worldwide (95% CI, 
9.2 to 36.7) could reduce their BMI in kg\(\cdot\)m\(^{-2}\) by participating 
in a regular exercise program.

For the four studies excluded because of insufficient 
data for BMI in kg\(\cdot\)m\(^{-2}\) [81–84], one reported a statistically 
significant exercise minus control group reduction (\(p = 0.02\)) in BMI in kg\(\cdot\)m\(^{-2}\) [84]. Another study that did not 
report results for BMI in kg\(\cdot\)m\(^{-2}\) did report a statistically 
significant exercise minus control group reduction in BMI 
\(z\)-score (\(p = 0.02\)) for the high-dose group as well as a 
trend for improvement (\(p = 0.06\)) in the low-dose group 
[81]. The remaining two studies did not report any BMI-
related results, although both reported statistically significant 
reductions of \(p < 0.01\) [82] and \(p < 0.001\) [83] for percent 
body fat. In addition, fail-safe \(N\) results indicated that a total 
of 774 studies with null findings would be needed to reverse 
our findings of a statistical significant reduction in BMI in 
kg\(\cdot\)m\(^{-2}\).

The results for trial sequential meta-analysis are shown 
in Figure 6. As can be seen, these findings confirm that the 
maximum information size has been reached and the stability 
of findings has been achieved with respect to exercise-
induced reductions in BMI in kg\(\cdot\)m\(^{-2}\) among overweight and 
obese children and adolescents. More specifically, changes 
in BMI in kg\(\cdot\)m\(^{-2}\) crossed the monitoring boundary for a 
type 1 error in 2010 and have remained stable thereafter. 
This confirms the statistical significance of exercise-induced
reductions in BMI in kg·m⁻² since 2010 among overweight and obese children and adolescents and suggests that the accumulation of additional studies in future years will not change these findings to one of nonsignificance. Simple metaregression results are shown in Supplementary File 5. No statistically significant associations were observed for those variables in which metaregression analysis was possible. Findings were similar when the one outlier for changes in BMI in kg·m⁻² was deleted from each of the analyses (results not shown) [66].

The results for GRADE with respect to changes in BMI in kg·m⁻² are shown in Supplementary File 6. Despite potential biases as well as heterogeneity, inconsistency, diversity, and overlapping prediction intervals, the overall quality of evidence was upgraded from low to moderate based on the magnitude of effect observed, trial sequential analysis results, and lack of adverse events.

### 3.3.2. Secondary Outcomes.

Secondary outcomes are shown in Table 3. Statistically significant improvements were found for body weight, fat mass, percent body fat, VO₂max in mL·kg⁻¹·min⁻¹, and upper and lower body strength. No statistically significant differences were observed for fat-free mass or energy intake. Changes were equivalent to relative improvements of 2.2% (body weight), 3.2% (fat mass), 2.9% (percent body fat), 10.3% (VO₂max in mL·kg⁻¹·min⁻¹), 8.9% (upper body strength), and 52.4% (lower body strength). With the exception of changes in upper and lower body strength, statistically significant heterogeneity as well as moderate to large inconsistency and diversity was observed for body weight, fat mass, percent body fat, and VO₂max in mL·kg⁻¹·min⁻¹. Prediction intervals were overlapping for all outcomes except for upper and lower body strength.

Statistically significant outliers (p < 0.05) were identified for changes in body weight [61, 66], fat mass [69], percent body fat [69], and VO₂max in mL·kg⁻¹·min⁻¹ [72]. With two outliers deleted from the model for body weight [61, 66], reductions remained statistically significant along with statistically significant heterogeneity, low inconsistency, and moderate diversity (X² = 13.9; 95% CI: 0 to 62.5; D² = 49.0%; 95% CI = 12.0 to 70.4). For fat mass, decreases remained statistically significant along with no statistically significant heterogeneity as well as very low inconsistency and diversity when one outlier was deleted from the model [69] (X² = 0.69; 95% CI: −0.17 to 1.22; z = −2.31; p < 0.001; Q = 13.9; I² = 0.31; I² = 13.9%; 95% CI = 0 to 62.5; D² = 21.1%). With the same study deleted [69], reductions in percent body fat remained statistically significant along with statistically significant heterogeneity, moderate inconsistency, and large diversity (X² = 13.9; 95% CI: 0.47 to 1.54; z = −3.68; p < 0.001; Q = 49.6; p < 0.001; I² = 69.7%; 95% CI = 49.5 to

---

**Figure 5:** Influence analysis results for point estimate changes in BMI in kg·m⁻² with each result deleted from the model once. The black squares represent the mean difference while the left and right extremes of the squares represent the corresponding 95% confidence intervals. The middle of the black diamond represents the overall mean difference while the left and right extremes of the diamond represent the corresponding 95% confidence intervals. Results are ordered from smallest to largest reductions.
Figure 6: Trial sequential analysis results. Trial sequential meta-analysis of exercise versus control for changes in BMI in kg m\(^{-2}\). The dashed inward sloping lines to the left represent trial sequential monitoring boundaries while the outward dashed sloping lines to the right represent futility boundaries. The solid black line represents the Z-curve and the black squares represent the cumulative results with each accumulating study from earliest (2004) to most recent (2014) year. The cumulative Z-curve, that is, black solid line with filled squares, crossed the monitoring boundaries in 2010, confirming that exercise reduces BMI in kg m\(^{-2}\) in overweight and obese children and adolescents and is unlikely to be reversed with additional studies in future years.

81.9; \(D^2 = 75.7\%\). Increases in \(\text{VO}_{2\text{max}}\) in mL kg\(^{-1}\) min\(^{-1}\) also remained statistically significant along with statistically significant heterogeneity, moderate inconsistency, and large diversity when one outlier [72] was deleted from the model (\(X^2: 2.35; 95\% \text{ CI}: 1.38 to 3.31; z = 4.76; p < 0.001; Q = 22.3; p = 0.004; I^2 = 64.1\%; 95\% \text{ CI}: 26.4 to 82.4; D^2 = 81.3\%\)). No outliers were identified for changes in lower and upper body muscular strength.

For those secondary outcomes in which statistically significant improvements were found, statistically significant small-study effects were observed for changes in percent body fat (\(\beta_0: 1.92, p = 0.03\)). No statistically significant small-study effects were observed for body weight (\(\beta_0: -0.54, p = 0.22\)), fat mass (\(\beta_0: -0.76, p = 0.26\)), \(\text{VO}_{2\text{max}}\) in mL kg\(^{-1}\) min\(^{-1}\) (\(\beta_0: -0.01, p = 0.50\)), or upper (\(\beta_0: 1.99, p = 0.15\)) and lower (\(\beta_0: 4.77, p = 0.27\)) body strength.

With each result deleted from the model once, changes remained statistically significant for all secondary outcomes in which the original findings were statistically significant. Changes ranged from -1.38 to -1.86 kg for body weight (25.3\% difference), -0.69 to -2.25 kg for fat mass (44.8\% difference), -1.00 to -1.53 for percent body fat (23.0\% difference), 2.34 to 3.42 mL kg\(^{-1}\) min\(^{-1}\) for \(\text{VO}_{2\text{max}}\) (31.6\% difference), 6.9 to 9.4 kg for upper body strength (26.6\% difference), and 41.5 to 46.0 kg for lower body strength (9.8\% difference).

4. Discussion

4.1. Findings. The overall findings of the current meta-analysis suggest that exercise is associated with reductions in BMI in kg m\(^{-2}\) among overweight and obese children and adolescents. Support for this interpretation is derived from (1) the overall magnitude of effect, (2) nonoverlapping 95\% CI, (3) continued significance when each study was deleted from the model once, including the one outlier [66], (4) apparent absence of small-study effects, (5) trial sequential analysis results demonstrating that the maximum information size had been reached and been stable since 2010, (6) the low NNT, and (7) the number of overweight and obese children and adolescents in the US and worldwide who might potentially improve their BMI in kg m\(^{-2}\) from the uptake of regular exercise. In addition, the magnitude of change in BMI in kg m\(^{-2}\) observed in this study (−1.08 kg m\(^{-2}\) or 3.6\%) may be clinically relevant as previous research has found significant improvements in selected health outcomes with a decrease in BMI in kg m\(^{-2}\) of approximately 4.8\% [85]. While the results of the current meta-analysis were 1.2\% smaller, they may still be clinically important. Regardless, the observed reductions in BMI in kg m\(^{-2}\) are most likely important at the population level. For example, a recent meta-analysis that reported a reduction of only 0.17 kg m\(^{-2}\) in BMI [20] as a result of school-based interventions suggested that their findings may result in important health benefits at the population level. This suggestion was based on the work of Rose [86] who contended that a small shift in population distribution can be an effective primary preventative strategy because more events occur among the large number of individuals at moderate risk than the small number at high risk. Importantly, the results of the current meta-analysis were more than six times larger than those of Lavelle et al. [20]. However, it is important to realize that whether an intervention should be recommended at the population level depends not only on the size of the effect but also on the costs associated with achieving such an effect as well as society’s willingness to pay for this. While the willingness of a society to pay for this most likely varies between countries and there is limited evidence regarding the cost-effectiveness of exercise interventions for the treatment of overweight and obesity in children and adolescents, one cost-effectiveness study found that the number of disability-adjusted life years was greater for a multifaceted school-based intervention that included physical education (8000) versus one without physical education (500) [87]. Clearly, further research in this area is needed.

Finally, considering that the results for GRADE were increased from low to moderate provides justification for recommending exercise for improving BMI in kg m\(^{-2}\) in overweight and obese children and adolescents. This is especially relevant given that a low rating is based on the belief that additional evidence in the future would most likely change the direction of effect, something that the investigative team does not believe will happen, especially given the trial sequential analysis results. In contrast to the investigative team’s findings that support the effects of exercise for reducing BMI in kg m\(^{-2}\) as
well as the fact that a random-effects model that incorporates heterogeneity into the analysis was used, no potential sources of heterogeneity were identified as a result of metaregression analyses. Thus, the current results could be compromised. This may be especially important given the large amount of inconsistency and diversity observed for BMI in kg·m$^{-2}$ in the current meta-analysis. However, while such analyses are important, covariate analyses in meta-analysis are considered observational given that studies are not randomly assigned to covariates [88]. As a result, such analyses do not support causal inferences [88]. Thus, while such analyses may generate important findings about potential sources of heterogeneity, they would still need to be tested in adequately powered randomized controlled trials [88]. A second finding that may weaken the BMI in kg·m$^{-2}$ results is the overlapping PI observed for changes in BMI in kg·m$^{-2}$. However, it is important to understand that PI are different compared to CI as the former are based on random-mean effects [35].

While no variables that accounted for heterogeneity with respect to changes in BMI in kg·m$^{-2}$ were found, it may be that factors that were unable to be assessed could account for some or all of the observed heterogeneity between the included studies. These include such things as (1) differences or changes in diet during the exercise intervention [89], (2) physical activity compensation [65, 90], and (3) genetic factors [91].

The results of the current meta-analysis are in agreement with one previous systematic review with meta-analysis that focused on exercise [20] but disagreement with four others that reported a nonsignificant decrease in BMI in kg·m$^{-2}$ among children and adolescents [17, 21–23]. Possible reasons for these discrepancies include (1) the small number of exercise-only studies that were included and pooled in these meta-analyses [17, 21, 22], (2) the inclusion of nonrandomized trials [20, 22], and (3) the inclusion of children and adolescents who were not overweight or obese [20, 22, 23].

The reductions in BMI in kg·m$^{-2}$ observed in the current meta-analysis also compare favorably to orlistat, the only weight-loss medication currently approved by the US Food and Drug Administration for the treatment of obese adolescents. In a recent meta-analysis, changes in BMI in kg·m$^{-2}$ that included two studies representing 579 participants resulted in a statistically significant decrease of $-0.76$ kg·m$^{-2}$ (95% CI, $-1.07, -0.44$) as a result of the use of orlistat [92]. These findings are approximately 30% less than those found for BMI in kg·m$^{-2}$ and exercise in the current meta-analysis.

The reductions in BMI in kg·m$^{-2}$ found in the current meta-analysis are also similar to the results reported in a recent systematic review of diet-only interventions in which decreases ranged from 0.8 to 2.7 kg·m$^{-2}$ [89]. This suggests that either exercise or diet can reduce BMI in kg·m$^{-2}$ in a similar fashion. In contrast, the results of this previous systematic review when diet and exercise were combined were equivocal, with changes in BMI in kg·m$^{-2}$ ranging from $-4.4$ to $0.27$ kg·m$^{-2}$ for aerobic exercise (4 studies), $-0.2$ to $1.1$ kg·m$^{-2}$ for resistance training (3 studies), and $-0.5$ to $-2.02$ kg·m$^{-2}$ for combined aerobic and resistance exercise (3 studies) [89]. However, whether these changes differ significantly according to type of exercise, type of diet, or some other factor(s) is not known.

In addition to changes in BMI in kg·m$^{-2}$, statistically significant and clinically important improvements in body weight, fat mass, percent body fat, relative VO$_{2\text{max}}$, and upper and lower body strength were observed. The changes in fat mass as well as percent body fat are particularly noteworthy since both are more relevant than BMI in kg·m$^{-2}$ with respect to improvements in body composition. However, because they are not as practical to assess, BMI in kg·m$^{-2}$ continues to be the preferred method of assessing and classifying overweight and obesity. In addition, the significant changes observed for the six secondary outcomes support the multiple benefits that can be derived from regular participation in exercise. The multiple benefits observed are in contrast to treatments such as pharmacological interventions, approaches that are usually intended to treat one outcome. In addition, orlistat, the only pharmacological intervention currently approved in the United States for the treatment of obesity in children and adolescents [93], has been shown to be less cost-effective than several nonpharmacologic interventions, including exercise [87, 94], and has also been accompanied by side-effects such as gastrointestinal distress [95]. With respect to exercise in the current meta-analysis, four studies that did include information on side-effects reported no serious adverse events [74, 77–79], defined as any intervention that results in death, a life threatening condition, hospitalization (initial or prolonged), disability, or permanent damage [96]. For these same four studies, adverse events, defined as any undesirable experience associated with an intervention, included primarily acute musculoskeletal injury or discomfort in 7.9% of exercise participants in one study [74] and none in the other three [77–79]. However, it is important to realize that 16 [61–73, 75, 76, 80] of the 20 studies in the current meta-analysis did not report adequate information with respect to adverse events.

4.2. Implications for Research. The results of this meta-analysis have several implications for both the reporting and conduct of future research. First, it is suggested that future studies report complete information regarding (1) allocation concealment, (2) blinding of outcome assessment, (3) dropouts according to each group, including reasons for dropping out, (4) adverse events, (5) the physical activity levels of participants prior to and during the intervention, (6) intensity of the exercise intervention, and (7) compliance to the exercise intervention. It is also suggested that investigators analyze and report results using both the per-protocol and intention-to-treat results. This will allow one to understand both the efficacy and effectiveness of exercise for improving BMI in kg·m$^{-2}$ as well as other outcomes in overweight and obese children and adolescents. In addition, since both energy intake and energy expenditure are critical in determining weight loss, future studies should collect and report data on energy intake and total daily energy expenditure. Finally, future studies should report complete
information on all outcomes assessed, partitioned by group. At a minimum, these data should include pre- and postsample sizes, means, and standard deviations as well as change outcome results along with their standard deviations.

It appears that a need exists for a four-arm randomized controlled trial in overweight and obese children and adolescents that includes an aerobic, strength, and combined aerobic and strength training group as well as a control group. Furthermore, to aid practitioners, a need exists for dose-response studies to determine the optimal exercise program(s) for overweight and obese children and adolescents. This may be especially important since it is currently recommended that children and adolescents participate in 60 or more minutes of moderate to vigorous physical activity per day (420 minutes per week) [97] but the current meta-analysis found statistically significant reductions in BMI in kg·m⁻² as well as several other outcome variables (body weight, fat mass, percent body fat, VO₂max, and upper and lower body strength) when the average total minutes per week was less than currently recommended. Finally, since cost is an important factor when deciding what intervention to recommend over another, a need exists for cost-effectiveness studies in overweight and obese children and adolescents.

4.4. Strengths and Potential Limitations of Current Study. In the investigative team’s opinion, there are at least four strengths of the current meta-analysis. First, to the best of the authors’ knowledge, this is the first trial sequential meta-analysis that has examined the effects of exercise on BMI in kg·m⁻² in overweight and obese children and adolescents, something that was not done in previous work by the investigative team [19]. This is important because it suggests that changes in BMI in kg·m⁻² are stable and not subject to a type 1 or type 2 error. Second, the inclusion of data regarding NNT, percentile improvement, relative improvement, and absolute number of overweight and obese children who might benefit from participation in a regular exercise program provides practical information to decision-makers with respect to what treatment, or combination of treatments, to recommend over others for overweight and obese children and adolescents. Third, the calculation of PI provides future researchers with an estimate of what effect they might expect to find for BMI in kg·m⁻² and several secondary outcomes (body weight, fat mass, percent body fat, VO₂max, and upper and lower muscular strength) if they were to conduct a randomized controlled exercise intervention trial in overweight and obese children and adolescents. Fourth, this supports previous work by the investigative team in which an exercise minus control group improvement of approximately 3% was found for BMI z-score [19]. From the investigative team’s perspective, the similar improvements observed for both BMI z-score and BMI in kg·m⁻² are important given the continued controversy regarding which metric is the most valid and reliable for assessing changes in adiposity among children and adolescents. For example, while one study reported that BMI z-score is the best BMI measure for assessing adiposity in children and/or adolescents [100], another [101], as well as more recent research [102], suggests that both absolute and relative changes in BMI in kg·m⁻² are better proxies for changes in adiposity. Thus, regardless of which BMI measure is superior for measuring changes in adiposity, something that is unlikely to be resolved in the near future, the investigative team’s previous [19] as well as current findings support similar exercise-induced improvements for both.

As opposed to the strengths of the current meta-analysis, there are at least five potential limitations. First, given the statistically significant heterogeneity as well as high inconsistency and diversity of the current findings as well as overlapping PI and GRADE findings, one might conclude that insufficient evidence currently exists to conclude that exercise is associated with statistically significant improvements in BMI in kg·m⁻² and selected secondary outcomes (body weight, fat mass, percent body fat, VO₂max in mL·kg⁻¹·min⁻¹, and

4.3. Implications for Practice. The results of the current meta-analysis suggest that exercise results in important improvements in BMI in kg·m⁻² as well as body weight, fat mass, percent body fat, VO₂max in mL·kg⁻¹·min⁻¹, and muscular strength in both upper and lower body. Lending further support for this contention is the low NNT, percentile improvement, and the estimated number of overweight and obese children in the United States and worldwide who could potentially benefit. Furthermore, no serious adverse events were reported for the four groups in which sufficient information was available. Unfortunately, the dose-response effects of exercise on BMI in kg·m⁻² and other outcomes in overweight and obese children and adolescents remain elusive. Thus, in order to not withhold a potentially beneficial and safe intervention and until more definitive evidence is available, it would appear prudent to recommend that practitioners follow the guidelines specific to children and adolescents as denoted in the 2008 Physical Activity Guidelines for Americans [98]. This includes at least 60 minutes per day of moderate to vigorous physical activity, primarily aerobic activity (running, hopping, skipping, jumping rope, swimming, dancing, and bicycling) as well as muscle strengthening activities and bone strengthening activities (running, jumping rope, basketball, tennis, hopscotch, etc.) [98].

While the focus of the current meta-analysis was on the effects of exercise on BMI in kg·m⁻² in overweight and obese children and adolescents, it would appear plausible to suggest that the addition of reduced caloric intake combined with exercise may result in even greater reductions in BMI in kg·m⁻². However, a recent meta-analysis of randomized controlled trials by Ho et al. found no statistically significant differences in BMI in kg·m⁻² between exercise and diet versus diet-only groups [89]. Importantly, the authors concluded that further randomized controlled trials with a rigorous design are needed to confirm their findings. Until that time, it would appear plausible to suggest that practitioners follow the recent recommendations that, in addition to exercise, include (1) the avoidance of sugar-sweetened beverages, (2) less recommendations that, in addition to exercise, include (1) the avoidance of sugar-sweetened beverages, (2) less high caloric density, and (3) increased intake of fruits and vegetables [99].
upper and lower muscular strength). Second, the statistically significant findings for increases in upper and lower body strength may need to be viewed with caution given that these findings were limited to three results. Consequently, the generalizability of these findings may be limited. Third, the generalizability of the current findings beyond the populations and intervention protocols included may be limited. Fourth, the results of the current meta-analysis, like any meta-analysis, may suffer from ecological fallacy, phenomena in which incorrect inferences about individual findings are made based upon aggregate statistics [103]. Fifth, since the search for studies focused on BMI in kg m\(^{-2}\) as the primary outcome, the results for all eight secondary outcomes may represent a biased sample.

5. Conclusions

The results of the current systematic review of previous meta-analyses suggest that exercise is associated with reductions in BMI in kg m\(^{-2}\) among overweight and obese children and adolescents. A need exists for randomized controlled trials to identify sources of heterogeneity, including dose-response studies.

Disclosure

Kristi S. Kelley and Dr. Russell R. Pate are the coauthors of this paper.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

The research reported in this paper was supported by the American Heart Association, Great Rivers Affiliate, Grant-in-Aid 12GRNT11670019 (GAK, Principal Investigator) and the National Institute of General Medical Sciences of the National Institutes of Health under Award no. U54GM104942. The content is solely the responsibility of the authors and does not necessarily represent the views of the American Heart Association or National Institutes of Health.

References


[97] Food and Drug Administration, What is a Serious Adverse Event? United States Food and Drug Administration, Silver Spring, Md, USA, 2014.


