Physical activity interventions and depression in children and adolescents: a systematic review and meta-analysis

Running title: PA interventions and depression

Authors: Helen E. Brown1, Natalie Pearson2, 4, Rock E. Braithwaite3, Wendy J. Brown1 and Stuart J. H. Biddle1, 2, 4

1School of Human Movement Studies, The University of Queensland, Brisbane, QLD, Australia.
2School of Sport, Exercise & Health Sciences, Loughborough University, Loughborough, Leicestershire LE11 3TU, UK.
3Department of Kinesiology and Recreation Administration, Humboldt State University, Arcata, CA, United States.
4 The NIHR Leicester-Loughborough Diet, Lifestyle and Physical Activity Biomedical Research Unit

Article type: Systematic review/meta-analysis

Keywords (not included in title): mental health, youth, exercise, programs

Abstract word count: 285 words

Manuscript word count: 2843 words (inclusive of all pages except the abstract and title page)

Abstract

Context: Evidence suggests chronic physical activity (PA) participation may be both protective against the onset of and beneficial for reducing depressive symptoms.

Objective: To assess the impact of PA interventions on depression in children and adolescents using meta-analysis.

Data sources: Published English language studies were located from manual and computerised searches of the following databases: PsycInfo, The Cochrane Database of Systematic Reviews and The Cochrane Central Register of Controlled Trials, Trials Register of Promoting Health Interventions (TRoPHI; EPPI Centre), Web of Science and MEDLINE.

Study selection: Studies meeting inclusion criteria (1) reported on interventions to promote or increase PA; (2) included children aged 5-11 years and/or adolescents aged 12-19 years; (3) reported on results using a quantitative measure of depression; (4) included a non-physical control or comparison group; (5) were published in peer-reviewed journals written in English, up to and including May 2011.

Data extraction: Studies were coded for methodological, participant, and study characteristics.

Comprehensive Meta-Analysis version-2 software was used to compute effect sizes; with sub-group

analyses to identify moderating characteristics. Study quality was assessed using the Delphi technique.

Results: Nine studies were included (N= 581); most were school-based RCTs randomized by individual. Studies used a variety of measurement tools to assess depressive symptoms. The summary treatment effect was small but significant (Hedges' g=-0.26, SE=0.09, 95% C.I = -0.43, -0.08, p=0.004). Sub-group analyses showed that methodological (e.g. studies with both education and PA intervention; those with a higher quality score; and less than three months in duration) and participant characteristics (e.g. single gender studies; those targeting overweight or obese groups) contributed most to the reduction in depression.

Conclusions: There was a small significant overall effect for PA on depression. More outcome-focused, high quality trials are required to effectively inform the implementation of programs to reduce depressive symptoms in children and adolescents.

Introduction

Current epidemiological data suggest a global prevalence of child and adolescent mental health disorders of approximately 20% (1) of which depression is the most frequently diagnosed (2). Depression during youth occurs frequently, is chronic and recurrent, and is associated with morbidity and mortality (3). Depression in children and adolescents can also result in substantial impairment in social functioning (such as relationships with peers and family members) (4), cognitive development and scholastic achievement (5) and associated suicide is reported to be the 3rd leading cause of death among adolescents (6).
Psychotherapy and pharmacological intervention are suggested as treatment options for depression in children and adolescents (3) but are often reported to be ineffective (7). Cognitive behavioural therapy can be of use but may be inaccessible and expensive (8). There is, however, growing interest in the use of physical activity (PA) interventions for the management of depression. Building on reviews that cite positive outcomes for the prevention and treatment of depression in adult populations (9-14), emerging evidence suggests similar outcomes may be attainable in children and adolescents.

A recent analysis (15) of the evidence from two systematic reviews of the effect of exercise on depression in this age group indicates encouraging results for the efficacy of PA interventions (7, 16). Statistically significant differences in depression scores for exercise conditions compared with no intervention are reported in a Cochrane Review by Larun and colleagues (7). Given the date of this review, it is timely to update the evidence. The purpose of this study, therefore, was to determine the overall efficacy of PA interventions on depression in children and adolescents by conducting a meta-analysis of available evidence published up to May 2011 and including both clinical and non-clinical populations. Both prevention (with non-clinical samples) and management (clinical) studies were included.

Methods

Search strategies and inclusion criteria

A broad literature search strategy was developed using keywords from five categories: population, study design, PA behaviour, mental health outcome and intervention type. Key terms (see Figure 1 for full search details) from each category were combined to locate all relevant literature using PsycInfo, The Cochrane Database of Systematic Reviews and The Cochrane Central Register of
Controlled Trials, Trials Register of Promoting Health Interventions (TRoPHI; EPPI Centre), Web of Science and MEDLINE. Relevant articles were selected by screening the titles and reviewing abstracts and, when abstracts were not available or did not provide sufficient data, the full text article was retrieved and screened to determine whether inclusion criteria were met. In addition, reference lists of primary studies, review papers, and identified articles were screened for titles that included key terms. Only studies that (1) reported measures on interventions to promote or increase PA; (2) included children and/or adolescents aged 5-19 years; (3) reported on results from any quantitative measure of depression; (4) included a non-physical control or comparison group; (5) were published in peer-reviewed journals written in English; and (6) up to and including May 2011 were included.

Data extraction

Information about study location and design, randomisation, setting, participants, target population, depression assessment, intervention, and effect sizes and their variation was extracted from each paper by two researchers independently. Study authors were emailed when missing information prevented data extraction. If author(s) did not respond to requests within one month, either the study or the relevant outcome was excluded from the analysis. Differences in facts (committed by transposing information during the coding process) and interpretation (coding differences based on interpretation of the available information) were discussed and resolved by the researchers.

Quality assessment

Study quality was assessed using the Delphi list(17). Studies were given scores (0 or 1) based on the method of randomisation, treatment allocation, similarity between groups at baseline, reporting of eligibility criteria, blinding of outcome assessors and participants, whether point estimates and measures of variability were presented for the primary outcome measures, and whether intention-
to-treat analysis was used. Each included paper was given a score out of eight with higher scores meaning higher study quality.

Effect Size Calculations

Comprehensive Meta-Analysis (CMA) version-2 software was used to compute all effect sizes in the current study(18). A random effects model using Hedges’ g as the effect size index was selected to measure differences in depression between PA experimental and comparison groups(19). The statistical assumption supporting a random effects model suggests that there will be within study error (sampling error) and between study variance. Standardised mean differences were adjusted by the inverse weight of the variance to prevent inflation of study weights, providing more accurate estimates of effect size. The meta-analytic literature has found that Hedges’ g prevents overestimation of an effect size value when sample sizes are fewer than 20 studies(20, 21). The standard formula for Hedges’ g used to correct for bias in small samples was:

\[
g = d \left[1 - \frac{\frac{3}{4N-9}}{\frac{1}{N}}\right]
\]

Where,

\[
d = \frac{M_1 - M_2}{S}
\]

In these equations, d represents Cohen’s (1969) formula for power(22); M is the sample mean, S the within group sample standard deviation, g is the effect size obtained from Hedges (1981) correction formula(23), using the total sample size N in the denominator.

Descriptive measures such as means, standard deviations, and sample sizes were used by CMA to calculate estimates of effect size. When descriptive data were insufficient then CMA provided options for effect sizes to be calculated from combinations of available data such as F, t, r, and/or p-values. When performing the analysis in CMA each study contributed one effect size calculation to this is a post-print version of the following article: Brown, Helen Elizabeth, Pearson, Natalie, Braithwaite, Rock E., Brown, Wendy J. and Biddle, Stuart J. H. (2013) Physical activity interventions and depression in children and adolescents: A systematic review and meta-analysis. Sports Medicine, 43 3: 195-206.
the overall analysis. When a study contained more than one measurement of depression, CMA provided an option that averaged outcomes to calculate an overall summary treatment effect.

Coding Studies were coded separately by two authors on three primary classifications (methodological characteristics, participant characteristics, and study characteristics), using processes that develop and refine coding sheets(24) to enable subgroup analysis. Methodological characteristics were coded according to research design (randomized control trial OR other design), intervention duration (< 3 months OR > 3 months), intervention frequency (< 3 days/week OR > 4 days/week), intervention delivery (Physical Activity, Education, OR Both), level of randomisation (school, class, OR individual), and study quality (Delphi score < 4 OR score > 5). Participant characteristics included sample size (N), participant age (mean age < 13 OR > 13), gender (males OR both males and females), and target population (Overweight, At-risk, OR School-based). Study characteristics were categorised by country (United Kingdom OR United States), study year (< 1996 OR > 1996), and measures (Depression only OR Multiple measures).

Heterogeneity of Variance The three statistics used to assess homogeneity of variance included the QTot al (QT) value which is based on a χ-square (χ2) distribution, tau-square (τ2) value, and I-square (I2) value. All three statistics (QT, τ2, and I2) were used to interpret heterogeneity of variance. When the QT statistic is significant then a procedure is used to conduct subgroup (moderator) analyses by compartmentalizing variance into QBetween (QB) and QWithin (QW) values, with significant QB values (p < .05) needing a statistical technique (T-test or ANOVA) to determine group differences(21). The τ2 statistic provides an estimate of total variance between studies with larger
values reflecting the proportion of variance that can be attributed to real differences between studies in a random effects model. When the number of studies per subgroup is small (k < 5) τ² can be imprecise, therefore, a pooled estimate of variance was used for all calculations(25). The I² statistic is the ratio of excess dispersion to total dispersion and can be interpreted as the overlap of confidence intervals explaining low (25%), moderate (50%), and high (75%) values of the total variance attributed to covariates(26). Larger values of I² require techniques (i.e., moderator analysis or meta-regression) to provide explanations(25, 26). Research suggests that smaller samples sizes increase the likelihood that assumptions will be violated when using a random effects model, as error can be overestimated(21). A conservative alpha level (α < .01) was established to prevent type I errors when interpreting results from the moderator analyses.

Outlier analysis and publication bias

Outliers were identified by analyzing relative residual values (Z < or > + 1.96) and if present were analyzed by using a “one study removed” technique that is available with the CMA version-2 software. The criterion for outlier inclusion was a large residual value that did not influence significant (p < .01) effect sizes (Hedges g) and remained within the 95% confidence interval. Publication bias was analyzed through visual inspection of a funnel plot, a Fail Safe N calculation(27) and a “Trim and Fill” procedure(28, 29). Funnel plots provide a visual representation of studies according to standard error (y-axis) and effect size (x-axis) with symmetrical distributions being indicative of a lack of publication bias. Fail Safe N calculations are based on the number of studies needed to nullify significant effects(27). The “Trim and Fill” procedure is an iterative statistical process that adds/removes studies to balance an asymmetrical funnel plot and provide an unbiased estimate of effect size(28, 29).

Results

The searches yielded 388 titles of potentially relevant articles; of which nine studies met the inclusion criteria (30-38). These studies comprised nine independent samples including 581 children and/or adolescents. The literature search strategy is shown in Figure 1, which also shows the primary reasons for exclusion of studies at each stage of the process.

FIGURE 1 NEAR HERE.

Methodological characteristics

Of the nine studies reviewed, five were randomised controlled trials (RCT) (30-32, 34, 37), two were controlled trials (CT) (35, 38), one was a cluster randomised controlled trial (33) and one was a quasi-experimental study (36) (see Table 1). Of these, four were randomised at the individual level (30-32, 37), three at the school level (34, 35, 38) and two at the class level (33, 36). Analysis for all studies was conducted at the individual level. Data used to compute effects sizes by CMA included independent group means and standard deviations for six papers (30, 32, 34, 35, 37, 38) and combinations of mean difference scores, t-values, and p-values in the three remaining studies (31, 33, 36).

Intervention periods ranged from nine (33) to 40 weeks (36) and sessions lasted between 20 (32) and 90 (31) minutes. Five studies held sessions on three days each week (30, 31, 36-38). The remaining studies held sessions on either two days (35), five days (32), four days (34) or 2-3 days (33). Intervention mode varied across studies. Aerobic exercise, included in six interventions, was the most common activity (30-32, 35, 37, 38). One intervention comprised health education (33), one was sport and physical education lessons (36), and one was yoga and mindfulness training (34). Three interventions were delivered by research staff (30, 32, 35), three by trained counsellors (31, 34, 38).
and two by physical education staff (36, 37). The remaining study did not provide delivery information (33).

TABLE 1 NEAR HERE.

Participant characteristics

Studies included between 19 (33) and 207 (32) participants (median sample 81 participants), and were conducted with both male and female participants. Two studies included only boys (31, 37) with the remaining studies using co-educational samples (30, 32, 33, 35, 36, 38). Participants were aged between 8 and 19 years.

The majority of studies targeted at-risk groups for depression. Two studies targeted clinically overweight young people (Body Mass Index Centile > 85th or 98th national average) (30, 37). Two studies included only criminally institutionalised youth offenders (31, 37). One study targeted those with low socio-economic status (36) and one those of Hispanic origin (33). The remaining three studies targeted the general population (34, 35, 38). Four interventions were conducted as part of the physical education curriculum (33-36), two in juvenile detention centres (31, 37) and two as part of larger interventions to reduce obesity (30) and improve metabolism (32). The remaining intervention was conducted in an after-school program (38).

Study characteristics

The majority of included interventions were conducted in the USA (31-34, 37, 38), with remaining interventions conducted in the UK (30, 35) and Chile (36). A variety of assessment tools was used to measure depressive symptoms. Three studies measured depression using the Beck Depression Inventory (BDI) (31, 33, 37). Remaining studies used the Children’s Depression Inventory (CDI) (30),
the Profile of Mood States (POMS)(38), the Hospital Anxiety and Depression Scale (HADS)(36), the Short Mood and Feelings Questionnaire (SMFQ-SF)(34), the Multiple Affect Adjective Check List (MAACL)(35) or the Reynolds Child Depression Scale (RCDS)(32).

Quality assessment

There was good initial agreement (89%) between the two independent reviewers on the methodological quality of included studies. Quality ratings for each study are shown in Table 2. Two studies scored 7 out of the maximum score of 8; neither blinded patients to the intervention(30, 32). A further study did not include an intention-to-treat analysis and did not blind participants and therefore scored 6 out of 8(37). The remaining studies received lower quality scores ranging from 5(33) to 2(35, 38).

Meta-analysis

Individual and summary effect size statistics are shown in Figure 2. When interpreting standardized mean differences, Cohen’s criteria(22) were used to determine small, r = .10; medium, r = .30; large, r = .50 effect sizes. Negative effect sizes were indicative of PA treatment groups having decreased depression scores when compared with control or comparison groups.

The PA intervention summary treatment effect for depression was small but significant (Hedges’ g = -0.26, SE = 0.09, 95% C.I. = -0.43, -0.08, p = 0.004) representing about one quarter of a standard deviation’s decrease in depression for experimental groups. Analysis revealed a significant heterogeneous distribution (QT = 19.84, p = .002, I² = 59.68) indicating the need to conduct subgroup (moderator) analyses using the methodological, participant, and study characteristics.
coded for each study. No outliers were identified as relative residuals (Z-values) ranged from 0.69 to -1.77, therefore, a sensitivity analysis was not conducted. The potential for publication bias was marginal as there was a symmetrical funnel plot, no studies were added when the Trim and Fill procedure was performed, and a moderate Fail Safe N calculation was needed to nullify a significant effect ($\alpha < 0.01$).

The review of homogeneity statistics found a moderate ($I^2 = 59.68$) significant ($Q_T = 19.84$) heterogeneous distribution that would explain a portion of the variance using subgroup analyses (see Tables 3 and 4). There were no significant ($p < .01$) differences between categories coded, however, trends were present with many subgroups. Borenstein and colleagues (25) recommend caution when interpreting subgroup analyses with fewer than five studies as estimates of effect size may be imprecise. We chose to report these findings only to provide guidance for future research by providing conservative interpretations of findings (25).

The trends from the moderator results suggest that methodological characteristics were important. The largest effects on depression were for higher quality, short (less than three months in duration), randomised controlled trials (randomised by individual) that included both education and physical activities. Participant characteristics that contributed most to the reduction of depression included single gender studies, studies that targeted overweight/obese youth or adolescence, and studies that were conducted in the United Kingdom. The only study characteristic that demonstrated a trend was interventions with a single measure of depression. All the PA intervention moderators demonstrated lower between study variance ($\tau^2$) and explained moderate to large portions of subgroup variance ($I^2$).
Discussion

This meta-analysis updated the extant literature on the impact of PA interventions on depression in children and adolescents. The results indicated a small but significant summary treatment effect. This is in contrast with the findings of Larun and colleagues (7) who conducted a systematic review of five exercise interventions on depression in children and adolescents (up to 20 years old), and found a significant moderate effect (standard mean difference effect size = −0.66; CI, −1.25, −0.08). They noted however that the number of included trials was low, that the trials were of low quality and highly varied in respect of methodological characteristics, such as sampling and measurement. We included results from more studies (nine, compared with five), including some considered to be of higher quality. As the results from the study by Norris and colleagues (which used flexibility activities and lower-intensity aerobic exercise) (35) contradicted those from all the other included studies, we considered the possibility that it’s inclusion may have contributed to our lower effect size. Sensitivity analysis found, however, that omission of this study had a marginal effect on the overall findings.

The results of individual studies in our meta-analysis were variable, as indicated by the moderate, significant heterogeneity statistics. This might be explained by our inclusion of both preventive and treatment trials, with a wide range of baseline depression levels, and with greater effects in the clinical trials. However, given the relative paucity of literature with children and adolescents, and the importance of mental health in this age group, it is sensible, in our view, not to be too restrictive with exclusion criteria at this point.

The findings also indicated that studies that focused on depression (i.e. did not measure additional outcomes) reported a greater treatment effect (30-33, 37). This suggests a possible need to test outcome-specific interventions, rather than generic ‘fix-all’ treatment approaches, when working with depression in children and adolescents. Moreover, young people referred to physical activity programmes for specific health reasons, for example as part of a study examining obesity (30) or metabolic outcomes (32), may have different motivations and expectation, which may affect psychological outcomes.

Our moderator analyses also demonstrated that studies with higher quality scores indicated greater treatment effects. These studies (30, 34) were distinct in that they concealed treatment allocation, blinded the outcome assessor and performed intention-to-treat analysis. Further experimental work should ensure high methodological quality using these criteria.

The majority of studies targeted at-risk groups for depression including criminally institutionalised male youth offenders (31, 37), those with low socio-economic status (36) and those of Hispanic origin (33). Two studies targeted young people with Body Mass Index (BMI) centile > 85th or 98th national average and reported a greater treatment effect of physical activity in clinically overweight participants (30, 32). Previous evidence also supports the premise that elevated BMI in children and adolescents is associated with depression (39, 40). Considering that PA is recognised as an effective strategy for weight loss (41, 42), such interventions may have dual benefit when working with overweight children and adolescents.

We are not able to do more than speculate on the likely mechanisms of why depression is reduced after physical activity. Plausible reasons include neuro-biological mechanisms, whereby neurotransmitters released during activity may mediate changes in depressive symptoms and mood.
and psychosocial mechanisms, which result in improved mastery and elevations in physical self-
worth, the latter being implicated in improved self-esteem(15, 43). However, there are no definitive
explanations as to which may be more important, and further work in this area is required. Coupled
with this is the need to investigate short and long term effects of physical activity, in the context of
possible spontaneous remission from depression with the passage of time.

Conclusion

This meta-analysis assessed the impact of PA interventions on depression in children and
adolescents. As only nine studies were included, caution is required in interpretation of the results.
However, study quality was higher than in previous reviews, and the small but significant treatment
effect suggests that physical activity may play a role in the prevention and treatment of depression
in young people. Further high quality outcome-focussed studies are now needed to confirm this
finding, and to inform the implementation of programs to effectively reduce depressive symptoms in
children and adolescents.

Funding sources: This work was supported by an Alf Howard International Travel Scholarship (School
of Human Movement Studies, University of Queensland) that enabled international collaborative
work to be undertaken. HEB was partially supported by an Australian NHMRC Program Grant (#
301110).

References


This is a post-print version of the following article: Brown, Helen Elizabeth, Pearson, Natalie,
and depression in children and adolescents: A systematic review and meta-analysis. Sports Medicine,


### Table 1: descriptive characteristics of studies meeting inclusion criteria

<table>
<thead>
<tr>
<th>Study (authors, country, study design)</th>
<th>Level of randomisation; number of clusters</th>
<th>Setting</th>
<th>Participants (number; age; % male)(^a)</th>
<th>Target population</th>
<th>Assessment period; depression measure</th>
<th>Intervention: description; delivery; duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annesi[1] (2005; USA; QE)</td>
<td>After-school program</td>
<td>After-school program</td>
<td>n=49; 10.5 ± 0.9yrs; 46%</td>
<td>None</td>
<td>0 and 12 weeks; POMS-SF</td>
<td>45 min aerobic, resistance and stretching exercise 3 day/week; delivered by trained counsellors; 12 weeks</td>
</tr>
<tr>
<td>Bonhauser et al.[2] (2005; Chile; CRT)</td>
<td>Class enrolment</td>
<td>Public school</td>
<td>n=198; 15.53 ± 0.9yrs; 54% additional education and 43% usual physical education</td>
<td>Low socio-economic status</td>
<td>0 and 40 weeks; HADS</td>
<td>90 min additional sport and physical education classes 3 day/week vs. usual physical education (comparison); delivered by physical education staff; 40 weeks</td>
</tr>
<tr>
<td>Daley et al.[3] (2009; UK; RCT)</td>
<td>Individual</td>
<td>Larger study population (examining the effect of exercise on obesity)</td>
<td>n=81; 13.1yrs; 44.4% BMI &gt;98(^{th}) percentile</td>
<td>0, 8, 14 and 28 weeks; CDI</td>
<td>30 min aerobic exercise 3 day/week; delivered by trained research staff; 28 weeks</td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Location</th>
<th>Setting</th>
<th>Participants</th>
<th>Intervention Details</th>
<th>Control Group Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hilyer et al.[4]</td>
<td>1982</td>
<td>USA; RCT</td>
<td>Individual</td>
<td>State industrial school for youth offenders; n=30; 15.5-18.6yrs; 100%</td>
<td>Labelled ‘criminally institutionalised’ 0 and 20 weeks; BDI</td>
<td>90 min ‘physical fitness’ program 3 day/week; delivered by counsellors; 20 weeks</td>
</tr>
<tr>
<td>MacMahon et al.[5]</td>
<td>1988</td>
<td>USA; RT</td>
<td>Individual</td>
<td>Juvenile detention centre; n=98; 16.3yrs (14-18.25); 100%</td>
<td>Labelled ‘criminally institutionalised’ 0 and 12 weeks; BDI</td>
<td>40 min vigorous exercise 3 day/week (intervention) vs. 40 min ‘light’ exercise 3 day/week (comparison); delivered by physical education staff; 12 weeks</td>
</tr>
<tr>
<td>Melnyk et al.[6]</td>
<td>2009</td>
<td>USA; CRCT</td>
<td>Class enrolment</td>
<td>Public school; n=19; 15.5 ± 0.63; 68%</td>
<td>Majority Hispanic 0 and 9 weeks; BDI</td>
<td>50 min health education (promoting physical activity) 2-3 day/week; delivery unknown; 9 weeks</td>
</tr>
<tr>
<td>Mendelson et al.[7]</td>
<td>2010</td>
<td>USA; CRCT</td>
<td>School enrolment</td>
<td>Public school; n=97; age range: 9.7-10.6; 39.2%</td>
<td>None 0 and 12 weeks; SMFQ-C</td>
<td>45 min yoga-based physical activity, breathing techniques and guided mindfulness practices 4 day/week; delivered by HLF instructors; 12 weeks</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Norris et al.[8] (1992; UK; CRCT)</th>
<th>School enrolment</th>
<th>Public school</th>
<th>n=60; mean age: 16.6yrs; 52%</th>
<th>None</th>
<th>0 and 10 weeks; Multiple Affect Adjective Check List</th>
</tr>
</thead>
<tbody>
<tr>
<td>Petty et al.[9] (2009; USA; RCT)</td>
<td>Individual</td>
<td>Larger study population (examining the effect of exercise on metabolism)</td>
<td>n=207; age range: 9.3-9.4yrs; 36% black and 52% white</td>
<td>BMI &gt;85th percentile</td>
<td>0 and 14 weeks; RCDS</td>
</tr>
</tbody>
</table>

Note. Study (authors, country, study design): USA = United States of America, UK = United Kingdom; QE = quasi-experimental study design, CRT = cluster-randomised trial, RCT = randomised controlled trial, CRCT = cluster-randomised controlled trial. Level of randomisation; number of clusters: N = number of clusters. Participants (number; age; %male): n = number of participants, yrs = years of age. Target population: BMI = Body Mass Index. Assessment period; depression measure: POMS-SF = Profile of Mood States–Short Form, HADS: Hospital Anxiety and Depression Scale, CDI = Children’s Depression Inventory, BDI = Beck Depression Inventory, SMFQ-C = Short Moods and Feelings Questionnaire-Children, RCDS: Reynolds Child Depression Scale. Intervention: description; delivery; duration: min = minutes, HLF = Holistic Life Foundation.

*aMeans and standard deviations and/or age ranges are provided where appropriate.

Table 2: quality assessment of studies meeting inclusion criteria

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Was a method of randomisation performed?</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Was the treatment allocation concealed?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Were the groups similar at baseline regarding the most important prognostic indicators?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Were the eligibility criteria specified?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Was the outcome assessor blinded?</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Was the patient blinded?</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Were point estimates and measures of variability presented for the primary outcome measures?</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Did the analysis include an intention-to-treat analysis?</td>
<td>n</td>
<td>Y</td>
<td>y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td><strong>Total Delphi score / 8</strong></td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Duration</th>
<th>Frequency</th>
<th>Delivery</th>
<th>Randomization</th>
<th>Quality</th>
<th>Age</th>
<th>Gender</th>
<th>Target</th>
<th>Country</th>
<th>Year</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annesi, 2005</td>
<td>O</td>
<td>&lt;3</td>
<td>1</td>
<td>PA</td>
<td>S</td>
<td>L</td>
<td>A</td>
<td>B</td>
<td>S</td>
<td>US</td>
<td>A</td>
<td>M</td>
</tr>
<tr>
<td>Bonhauser et al., 2005</td>
<td>O</td>
<td>≥3</td>
<td>1</td>
<td>PA</td>
<td>C</td>
<td>L</td>
<td>A</td>
<td>B</td>
<td>S</td>
<td>US</td>
<td>A</td>
<td>M</td>
</tr>
<tr>
<td>Daley et al., 2006</td>
<td>RCT</td>
<td>&lt;3</td>
<td>1</td>
<td>PA</td>
<td>I</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>O</td>
<td>UK</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>Hilyer et al., 1982</td>
<td>RCT</td>
<td>≥3</td>
<td>1</td>
<td>B</td>
<td>I</td>
<td>L</td>
<td>A</td>
<td>M</td>
<td>A</td>
<td>US</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>McMahon &amp; Gross, 1988</td>
<td>RCT</td>
<td>≥3</td>
<td>1</td>
<td>PA</td>
<td>I</td>
<td>H</td>
<td>A</td>
<td>M</td>
<td>A</td>
<td>US</td>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>Melnyk et al., 2009</td>
<td>O</td>
<td>≥3</td>
<td>1</td>
<td>E</td>
<td>S</td>
<td>H</td>
<td>A</td>
<td>B</td>
<td>S</td>
<td>US</td>
<td>A</td>
<td>D</td>
</tr>
<tr>
<td>Mendelson et al., 2010</td>
<td>RCT</td>
<td>&lt;3</td>
<td>2</td>
<td>B</td>
<td>C</td>
<td>L</td>
<td>Y</td>
<td>B</td>
<td>S</td>
<td>US</td>
<td>A</td>
<td>M</td>
</tr>
<tr>
<td>Norris et al., 1992</td>
<td>O</td>
<td>&lt;3</td>
<td>1</td>
<td>PA</td>
<td>C</td>
<td>L</td>
<td>A</td>
<td>B</td>
<td>S</td>
<td>UK</td>
<td>B</td>
<td>M</td>
</tr>
<tr>
<td>Petty et al., 2009</td>
<td>RCT</td>
<td>≥3</td>
<td>2</td>
<td>PA</td>
<td>I</td>
<td>H</td>
<td>Y</td>
<td>B</td>
<td>O</td>
<td>US</td>
<td>A</td>
<td>D</td>
</tr>
</tbody>
</table>

Note. Design (Research Design): O = other design, RCT = randomized control trial. Duration (Research Duration): < 3 = less than 3 months, ≥ 3 = greater than or equal to 3 months. Frequency (Intervention Frequency): 1 = less or equal to 3 days/week, 2 = greater than 4 days per week. Delivery (Intervention Delivery): PA = physical activity, E = education, B = both PA and education. Randomization (Randomization Level): S = School, C = class, I = individual. Quality (Study Quality): L = Delphi score < 4, H = Delphi score > 5. Age: A = adolescent mean age > 13 years, Y = youth mean age < 13 years. Gender: M = males only, B = Both males and females. Target (Intervention Target): S = school based, O = obese/overweight, A = at-risk/adolescent.
Table 4: sub-group analysis of included studies
Effect Sizes (95% Confidence Limits) and Between-Group Tests of Heterogeneity ($Q$), Assessment of the Amount of Variability Caused by Random Error ($I^2$), and Tests of Moderators for Studies Included in the Meta-analysis

<table>
<thead>
<tr>
<th>Methodological Characteristics</th>
<th>Random Effects Model $^a$</th>
<th>Effect Size Statistics</th>
<th>Null Test</th>
<th>Heterogeneity Statistics</th>
<th>Publication Bias</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$g$</td>
<td>SE</td>
<td>$s^2$</td>
<td>95% C.I.</td>
</tr>
<tr>
<td>Research Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Randomized Control Trial</td>
<td>5</td>
<td>-0.345</td>
<td>0.110</td>
<td>0.012</td>
<td>(-0.561, -0.129)</td>
</tr>
<tr>
<td>Other Design</td>
<td>4</td>
<td>-0.135</td>
<td>0.132</td>
<td>0.017</td>
<td>(-0.393, 0.123)</td>
</tr>
<tr>
<td>Intervention Duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 3 Months</td>
<td>6</td>
<td>-0.329</td>
<td>0.108</td>
<td>0.012</td>
<td>(-0.541, -0.118)</td>
</tr>
<tr>
<td>≥ 3 Months</td>
<td>3</td>
<td>-0.149</td>
<td>0.135</td>
<td>0.018</td>
<td>(-0.413, 0.115)</td>
</tr>
<tr>
<td>Intervention Frequency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 3 Days/Week</td>
<td>7</td>
<td>-0.288</td>
<td>0.107</td>
<td>0.011</td>
<td>(-0.497, -0.078)</td>
</tr>
<tr>
<td>&gt; 4 Days/Week</td>
<td>2</td>
<td>-0.156</td>
<td>0.190</td>
<td>0.036</td>
<td>(-0.528, 0.216)</td>
</tr>
<tr>
<td>Intervention Delivery</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>6</td>
<td>-0.198</td>
<td>0.096</td>
<td>0.009</td>
<td>(-0.386, -0.009)</td>
</tr>
<tr>
<td>Both PA &amp; E</td>
<td>2</td>
<td>-0.458</td>
<td>0.147</td>
<td>0.022</td>
<td>(-0.746, -0.170)</td>
</tr>
<tr>
<td>Level of Randomization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School</td>
<td>2</td>
<td>-0.275</td>
<td>0.197</td>
<td>0.030</td>
<td>(-0.661, 0.110)</td>
</tr>
<tr>
<td>Class</td>
<td>3</td>
<td>-0.061</td>
<td>0.142</td>
<td>0.020</td>
<td>(-0.611, -0.161)</td>
</tr>
<tr>
<td>Individual</td>
<td>4</td>
<td>-0.386</td>
<td>0.115</td>
<td>0.013</td>
<td>(-0.433, 0.080)</td>
</tr>
<tr>
<td>Study Quality</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low-Score &lt; 4</td>
<td>5</td>
<td>-0.145</td>
<td>0.111</td>
<td>0.012</td>
<td>(-0.363, 0.074)</td>
</tr>
<tr>
<td>High-Score &gt; 5</td>
<td>4</td>
<td>-0.386</td>
<td>0.116</td>
<td>0.013</td>
<td>(-0.614, -0.159)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>&lt; 13 years</th>
<th>≥ 13 Years</th>
<th>Males Only</th>
<th>Males &amp; Females</th>
<th>Obese/Overweight</th>
<th>At-Risk</th>
<th>School</th>
<th>UK</th>
<th>US</th>
<th>Country</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>-0.156</td>
<td>-0.288</td>
<td>-0.351</td>
<td>-0.233</td>
<td>-0.407</td>
<td>-0.249</td>
<td>-0.143</td>
<td>-0.358</td>
<td>-0.223</td>
<td>-0.358</td>
<td>-0.233</td>
<td>-0.223</td>
</tr>
<tr>
<td></td>
<td>0.190</td>
<td>0.107</td>
<td>0.218</td>
<td>0.106</td>
<td>0.162</td>
<td>0.158</td>
<td>0.146</td>
<td>0.165</td>
<td>0.099</td>
<td>0.146</td>
<td>0.146</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>0.036</td>
<td>0.011</td>
<td>0.048</td>
<td>0.011</td>
<td>0.026</td>
<td>0.025</td>
<td>0.021</td>
<td>0.027</td>
<td>0.010</td>
<td>0.021</td>
<td>0.021</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(-0.528, 0.216)</td>
<td>(-0.497, -0.078)</td>
<td>(-0.778, 0.077)</td>
<td>(-0.440, -0.026)</td>
<td>(-0.725, -0.089)</td>
<td>(-0.559, 0.062)</td>
<td>(-0.454, -0.064)</td>
<td>(-0.682, -0.034)</td>
<td>(-0.417, -0.029)</td>
<td>(-0.682, -0.034)</td>
<td>(-0.417, -0.029)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.058</td>
<td>0.042</td>
<td>0.049</td>
<td>0.876</td>
<td>0.329</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.219</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Study Characteristics**

<table>
<thead>
<tr>
<th>Year</th>
<th>&lt; 13 years</th>
<th>≥ 13 Years</th>
<th>Males Only</th>
<th>Males &amp; Females</th>
<th>Obese/Overweight</th>
<th>At-Risk</th>
<th>School</th>
<th>UK</th>
<th>US</th>
<th>Country</th>
<th>UK</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before 1996</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>1996 and After</td>
<td>-0.200</td>
<td>-0.279</td>
<td>-0.245</td>
<td>-0.279</td>
<td>-0.407</td>
<td>-0.249</td>
<td>-0.143</td>
<td>-0.358</td>
<td>-0.223</td>
<td>-0.358</td>
<td>-0.233</td>
<td>-0.223</td>
</tr>
<tr>
<td></td>
<td>0.172</td>
<td>0.110</td>
<td>0.111</td>
<td>0.110</td>
<td>0.162</td>
<td>0.158</td>
<td>0.146</td>
<td>0.165</td>
<td>0.099</td>
<td>0.146</td>
<td>0.146</td>
<td>0.099</td>
</tr>
<tr>
<td></td>
<td>0.029</td>
<td>0.012</td>
<td>0.012</td>
<td>0.012</td>
<td>0.026</td>
<td>0.025</td>
<td>0.021</td>
<td>0.027</td>
<td>0.010</td>
<td>0.021</td>
<td>0.021</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(-0.536, 0.136)</td>
<td>(-0.495, -0.063)</td>
<td>(-0.563, -0.127)</td>
<td>(-0.454, -0.064)</td>
<td>(-0.725, -0.089)</td>
<td>(-0.559, 0.062)</td>
<td>(-0.417, -0.029)</td>
<td>(-0.682, -0.034)</td>
<td>(-0.417, -0.029)</td>
<td>(-0.682, -0.034)</td>
<td>(-0.417, -0.029)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.043</td>
<td>0.054</td>
<td>0.050</td>
<td>0.001</td>
<td>0.042</td>
<td>0.011</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**Note:** PA = physical activity, E = education.  
\(k\) = number of effect sizes.  
\(g\) = effect size (Hedges g).  
SE = standard error.  
\(S^2\) = variance.  
CI = confidence intervals (lower limit, upper limit).  
Z = test of null hypothesis.  
\(\tau^2\) = between study variance in random effects model.  
\(I^2\) = total variance explained by moderator.  * indicates \(p < .01\).  
\(Q\) = Total Q-value used to determine heterogeneity.  
\(Q\) = Between Q-value used to determine significance (\(a < .01\)).