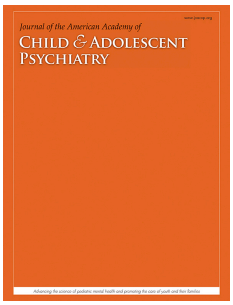
Accepted Manuscript



The Effect of Exercise on Depressive Symptoms in Adolescents: A Systematic Review and Meta-Analysis

Tim Carter, PhD, Ioannis D. Morres, MSc, Oonagh Meade, PhD, Patrick Callaghan, PhD

PII: S0890-8567(16)30170-8

DOI: 10.1016/j.jaac.2016.04.016

Reference: JAAC 1428

To appear in: *Journal of the American Academy of Child & Adolescent Psychiatry*

Received Date: 10 November 2015

Revised Date: 16 February 2016

Accepted Date: 30 April 2016

Please cite this article as: Carter T, Morres ID, Meade O, Callaghan P, The Effect of Exercise on Depressive Symptoms in Adolescents: A Systematic Review and Meta-Analysis, *Journal of the American Academy of Child & Adolescent Psychiatry* (2016), doi: 10.1016/j.jaac.2016.04.016.

The Effect of Exercise on Depressive Symptoms in Adolescents: A Systematic Review and Meta-Analysis

RH: Exercise for Depression in Adolescents

Tim Carter, PhD; Ioannis D. Morres, MSc; Oonagh Meade, PhD; Patrick Callaghan, PhD

This article is discussed in an editorial by Dr. Peter Jensen on page xx.

Clinical guidance is available at the end of this article.

Accepted May 3, 2016

Drs. Carter, Meade, and Callaghan are with University of Nottingham, United Kingdom. Mr. Morres is with University of Thessaly, Greece.

Dr. Carter served as the statistical expert for this research.

The authors would like to acknowledge Boliang Guo, PhD, medical statistician, University of Nottingham, for his expert statistical advice.

Disclosure: Drs. Carter, Meade, Callaghan, and Mr. Morres report no biomedical financial interests or potential conflicts of interest.

Correspondence to Tim Carter, PhD, Room D25, Institute of Mental Health, Triumph Road, Nottingham, NG7 2TU, United Kingdom; email: Timothy.carter@nottingham.ac.uk.

Abstract

Objective: The purpose of this review was to examine the treatment effect of physical exercise on depressive symptoms for adolescents aged 13-17 years. **Method:** A systematic search of seven electronic databases identified relevant randomised controlled trials. Following removal of duplicates, 543 texts were screened for eligibility. Screening, data extraction, and trial methodological quality assessment (using the Delphi list) were undertaken by two independent researchers. Standardized mean differences were used for pooling post-intervention depressive symptom scores. **Results:** Eleven trials met the inclusion criteria, eight of which provided the necessary data for calculation of standardized effect size. Exercise showed a statistically significant moderate overall effect on depressive symptom reduction (SMD -0.48, 95% CI -0.87, -0.10, p= .01, I^2=67%). Amongst trials with higher methodological scoring, a non-significant moderate effect was recorded (SMD -0.41, 95% CI -0.86, 0.05, p= .08). In trials with exclusively clinical samples, exercise showed a statistically significant moderate effect on depressive symptoms with lower levels of heterogeneity (SMD - 0.43, 95% CI -0.84, -0.02, p= .04), I^2=44%). **Conclusion:** Physical exercise appears to improve depressive symptoms in adolescents, especially in clinical samples where the moderate antidepressant effect, higher methodological quality, and lowered statistical heterogeneity suggest exercise may be a useful treatment strategy for depression. Larger trials with clinical samples that adequately minimise the risk of bias are required for firmer conclusions on the effectiveness of exercise as an antidepressant treatment. **Key words:** Exercise, randomised controlled trial, depression, adolescents, meta-analysis

Introduction

The lifetime prevalence of depression in adolescents has doubled between the mid-1980s and 2000s.[1] To this extent, the prevalence of major depressive disorder (MDD) is now ranging from 4% to 8%, and as many as 12% of children and adolescents may have subthreshold symptoms of depression.[2] Alarmingly, 20% of young people experience at least one episode of major depression before they reach 18 years of age.[3]

In recent years, there has been an increased, yet limited, number of studies investigating the effect of physical exercise or physical activity for depression in children and adolescents (physical exercise is defined as a planned, structured, repetitive, and purposive physical activity, aiming to improve or maintain one or more physical fitness components) [4], and there is some evidence suggesting this may be a promising strategy [5,6]. In particular, two previous systematic reviews have attempted to synthesise these studies: Larun et al[5] reported a small to moderate treatment effect in favour of exercise, and Brown et al[6] found a small treatment effect for physical activity over nonphysical activity comparisons. The conclusions from these reviews, however, are not without limitations. Larun et al[5] reported that the majority of trials tended to have small sample sizes and were of low methodological quality, while Brown et al[6] included quasi-experimental studies alongside randomised controlled trials (RCTs). Also, both reviews covered a wide age range of participants (8-20 years of age). Considering the substantial biological and psychosocial changes taking place during childhood and adolescence, the effect of exercise may be different depending on age-related developmental levels. As such, it is likely this may have had a significant confounding impact on previous pooled findings. Considering also that depression is more prevalent among adolescents than children,[7-9] alongside the recent increased attention to this field, a meta-analytic review is warranted to examine the antidepressant effect of exercise in

adolescents. This review will specifically explore the effect of exercise interventions on depressive symptoms for adolescents aged 13-17 years.

Method

Search strategy

A broad literature search strategy was developed using keywords, Medical Subject Headings (MeSH), and title search terms in four categories: population, intervention type, outcomes, and study design. The search terms were based on modified versions of the terms used in the previous systematic reviews in this field[5,6] and were implemented with assistance from an information specialist in the Greenfield Medical library at the University of Nottingham, United Kingdom.

The search terms from each category were combined to locate all relevant literature using the following databases: The Cochrane Library; MEDLINE; EMBASE; CINAHL; Sportdiscus; AMED, and PsychInfo. The clinicaltrials.gov registry was also searched to identify trials that were not yet published but had relevant outcome data available. Slight variations in MeSH terms and keywords were used depending on the database being searched (see Table 1 for a full list of search terms used to search the EMBASE database). The search was not restricted by date. The search was last updated in April 2014.

INSERT TABLE 1 HERE

Inclusion Criteria

Participants

Study participants had a mean/median age between 13 and 17 years.

Interventions

Studies reporting an intervention that i) promoted exercise or physical activity; ii) identified specified session duration and an overall time period or total number of sessions; iii) was implemented as a monotherapy or complemented by routine physical activity, an educational

component, or treatment as usual (however defined); and iv) was conducted supervised or unsupervised, individually based or in a group setting were included.

Comparison

Studies that included at least one of the following comparison conditions—no treatment, waiting list, treatment as usual (however defined), a psychosocial intervention, an educational intervention, an equivalent contact (attention) stretching condition—were included.

Outcomes

Studies that reported findings from a continuous, validated outcome measure of depressive symptoms were included. The outcome measure could be completed via self-report, researcher, teacher, parent, or clinician.

Study design

Only RCTs were included in order to limit the potential biasing effects of including studies with greater risks to internal validity. Studies using either individual or cluster randomisation were included. No restriction was placed on publication date or geographical location of the studies.

Exclusion Criteria

Participants

Exclusion criteria included a mean/median participant age outside the range of 13-17 years; target populations with identified physical health problems (e.g. obesity, diabetes, cancer); intellectual disabilities; or eating disorders.

Intervention

Studies with no inclusion of a physical activity/exercise intervention (however defined) or no information on the duration of the intervention were excluded.

Outcomes

Studies with no validated, continuous measure of depressive symptoms (however defined) were excluded.

Study Design

Non-RCTs, RCTs that were not available in the English language, and RCTs not published in peer reviewed journals were not included.

Screening

Two reviewers (T.C. and O.M.) independently screened titles and abstracts of all identified articles following removal of duplicates. 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 review papers and identified articles were screened for titles that included key terms. Articles were excluded using the pre-specified exclusion criteria. Where data were obtained for trials not yet published, such data were included in the meta-analysis, but the trials were not included in the quality appraisal or qualitative synthesis. However, where such trials were published following this period, they were fully incorporated into the review findings.

Data Collection and Analysis

Primary Outcome

The primary outcome was depressive symptoms as measured via a continuous scale. In order to avoid applying parametric tests to nonparametric data, the following standards were applied to the endpoint data derived from the continuous depression rating scales: (a) standard deviations (SD) and means had to be obtainable for endpoint measures on the rating scales used; (b) the SD when multiplied by two had to be less than the mean (otherwise the mean was considered skewed and unlikely to be an appropriate measure of central tendency).[10]

Measures of Treatment Effect

For the primary outcome, standardised mean differences (SMD) were calculated between groups. It would have been preferential not to calculate effect size measures (SMD), as the output is more meaningful when presented as a mean difference (MD). However, based on the previous reviews in this area, it was expected that the depression scales would vary between studies, and therefore the SMD was deemed appropriate.[11]

Quality Assessment

Two reviewers (T.C. and I.M.) independently assessed the methodological quality of the included studies using the Delphi list.[12] The Delphi list is a nine-item checklist developed for reporting quality scores of RCTs. For the purpose of this review, one item was removed from the Delphi List: blinding of therapist, as this is not feasible in trials of exercise because the nature of the intervention does not allow for therapist blinding. The Delphi list in this review therefore allowed for scores (0 or 1) for each of the following eight criteria: method of randomization; 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.

A score of 0 indicates that the item under assessment was not reported, whereas a score of 1 indicates that the item under assessment was reported in sufficient detail. In instances where there was partial reporting of an item, or only a suggestion that it may have been undertaken, then it was coded as unclear, and subsequently scored as 0. Each assessed study was given a score out of eight, with higher scores signifying higher study quality. Following comparison on initial independent scoring, all disagreements were resolved through discussion, and agreed quality scores were produced. For the purpose of the sensitivity analysis, studies scoring five and above (≥5) were considered studies with higher-

quality scoring (this score represents over half the Delphi List items having been sufficiently fulfilled and reported).

Assessment of heterogeneity and sensitivity analysis

Initially, a visual inspection of the forest plots was undertaken to investigate the possibility of statistical heterogeneity. Following this, heterogeneity between studies was determined through observation of the I^2 value. The I^2 provides an estimate of the percentage of inconsistency thought to be due to chance.[11] An I^2 estimate greater than or equal to 50% was interpreted as evidence of substantial levels of heterogeneity.[11] When high levels of heterogeneity were found in the primary outcome, sensitivity analysis was undertaken to explore possible causes of the heterogeneity.

Four sensitivity analyses were planned prior to the review: (1) trials with higher-quality scores on the Delphi list to investigate whether increased risk of bias inflated the treatment effect; (2) the removal of obvious outliers on the forest plot in terms of effect size; (3) trials with participants in treatment for a depressive related disorder or with a diagnosis of a depressive disorder only; (4) trials with participants from the general population only (non-clinical).

Importantly, based on previous systemic reviews in this area, it was expected that any sensitivity analyses would be conducted with small numbers of trials. Therefore, caution was taken when interpreting the sensitivity analyses, as it is suggested that fewer than five studies as estimates of effect size may be imprecise.[13]

Data synthesis

Review Manager software (version 5.2) was used for statistical pooling. A random-effects model was used when pooling the data to reduce the bias stemming from the potential heterogeneity between studies. Where studies included more than one intervention arm, they were analysed separately in the analysis.

Data extraction

Data extraction was independently undertaken by two researchers (T.C. and O.M.) using a modified version of the Cochrane Effective Practice and Organization of Care Group data extraction form.[14] All inconsistencies were resolved through discussion. Where descriptive data for the primary outcome was not available, authors were contacted.

Results

Following the removal of duplicates, 543 potentially relevant articles were identified. Eleven trials met the inclusion criteria and were comprised of 11 independent samples (1,449 participants), with eight trials eligible for meta-analysis. The primary reason for exclusion at each stage of the process is given in the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flow diagram[15] in Figure 1. This was created using Review Manager software (version 5.2).

INSERT FIGURE 1 HERE

Description of Study Design

A full description of the characteristics of each included trial is given in Table 2. All trials were published between 1982 and 2015 and were reported in the English language. Seven trials were conducted in the United States, [16-22] one trial was conducted in Chile,[23] one was conducted in Iran,[24] one was conducted in South Korea,[25] and one was conducted in the United Kingdom.[26]

Seven trials allocated participants through individual randomisation,[16,17,20,22,24-26] and four reported using cluster randomisation.[18,19,21,23]

Description of Study Participants

Sample sizes ranged between 19 and 779 participants (median=60). Two trials included females only,[24,25] one trial included males only,[20] and all remaining trials included both genders. The mean age of participants in each trial ranged from 14.7 to 17 years of age.

Five trials included participants who were volunteers from the general population, all of which were high school students.[18,19,21,23,25] One trial included participants with scores indicating moderate depression from an "at risk" population in a juvenile delinquent institution.[20] Five trials recruited clinical samples: two trials recruited inpatients receiving treatment for depression or dysthymia from psychiatric inpatient settings[16,22]; two trials included participants with a diagnosis of depression who were not currently receiving psychiatric treatment[17,24]; and one trial included participants who were receiving outpatient treatment for depression.[26] Based on the average baseline depression score, severity of depression was interpreted as follows: severe in two trials[24,26]; moderate to severe in one trial [22]; moderate in two trials[17,20]; mild in two trials[16,25]; average (i.e., below mild) in three trials [19,21,23]; and unable to be classified in one trial.[18]

Only two trials included baseline exercise participation data,[25,26] reporting that participants were sedentary. No trial reported targeting adolescents who engaged in above-average levels of exercise or who regularly exercised.

Description of Outcome Measures

A range of outcome measures was used to quantify depressive symptoms. All trials utilized self-report outcome measures. Five trials used an outcome measure that focused solely on depression: the Beck Depression Inventory (BDI)[16,20]; the Children's Depression Rating Scale–Revised (CDRS-R)[17]; the Children's Depression Inventory (CDI)[22]; the Children's Depression Inventory 2nd Version (CDI-2)[26]; and the Hamilton Depression Rating Scale. [24]

Three trials used outcome measures that captured both anxiety and depressive symptoms: the Beck Youth Inventory (BYI),[19,21] and the Hospital Anxiety and Depression Scale (HADS).[23]

Two trials used an outcome measure within which depressive symptoms were measured through a subscale: Behaviour Assessment System for Children, Second Edition (BASC-2),[18] and the Symptom Check List-90-Revision (SCL-90-R).[25]

Description of Intervention and Comparison Conditions

In all trials, physical exercise was employed as a singular or complementary intervention in the experimental group and included some form of aerobic and/or resistance/strength training. The intervention was conducted within school settings in five trials, [18,19,21,23,25] within institution or inpatient setting in three trials,[16,20,22] and within community-based settings in three trials.[17,24,26]

A limited number of trials provided an adequate description of the employed intensity exercise. In particular, three trials prescribed set exercise intensities for participants, [17,22,24] and one trial promoted preferred intensity exercise. [26] These four trials recruited clinical samples. The remaining trials, however, did not give any information regarding exercise intensity. In two trials, an educational component accompanied the exercise intervention.[19,21] All the trials implemented supervised group exercise interventions. Three trials encouraged additional unsupervised exercise between supervised sessions. [17,19,21]

Intervention duration ranged from six weeks [24,26] to 40 weeks.[23] The median duration across all of the trials was 11 weeks, and the median duration across the trials with clinical samples was seven weeks. The majority of the trials implemented the exercise intervention three times per week.[16,17,20,22-25]

Four trials employed usual exercise routine (as prescribed by the participants' school or institution) as a control.[16,18,20,23] Four trials used equivalent contact conditions between intervention and control groups.[17,19,21,22] Two trials used a no-treatment control condition,[24,25] and one trial used psychiatric treatment as usual.[26] The majority of trials did not report the participant's baseline exercise participation or their concurrent exercise participation during

the intervention period (in addition to the implemented exercise in the intervention and control groups). No trials reported targeting participants with high levels of physical exercise or habitual physical activity. For trials with samples from the general population (high school students), it is assumed that participants took part in concurrent normative physical education.

Quality Assessment of Included Studies

The interrater reliability coefficient for the two independent ratings on the Delphi list was high (Kappa = 0.81, 95% CI 0.69, 0.94), indicating strong initial agreement between reviewers on assessment of methodological quality. [27,28] Consensus scoring for each of the included trials is provided in Table 3. One trial scored seven out of a maximum of eight on the Delphi list;[26] one trial scored six;[25] four trials scored five; [3,16,19,27] the remainder scored four and below, with one trial scoring one.[16]

Three trials reported using intention to treat (ITT) analysis.[21,23,26] Two trials reported blinding outcome assessors to treatment condition.[17,26] One trial adequately described allocation concealment procedures.[26]

Reporting Bias

A visual inspection of the funnel plot for the primary outcome was undertaken, and there appeared to be no evidence of asymmetry.

INSERT TABLE 2 HERE

INSERT TABLE 3 HERE

Meta-Synthesis of Primary Outcome

Three trials could not be combined in the meta-analysis: one trial failed to report the standard deviation for the experimental group and did not include sufficient data to estimate it;[16] one trial did not report the standard deviation of post-intervention outcomes; [29] and one did not report means or standard deviations at post intervention.[21]

Individual and overall effect size statistics are shown in Figure 2. Negative effect sizes were indicative of the exercise intervention groups having decreased depression scores when compared with control or comparison groups.

Exercise showed an overall moderate, statistically significant treatment effect on depressive symptoms (SMD [random effects] -0.48, 95% CI -0.87, -0.10, p= .01). Heterogeneity observed in this analysis was moderate (I^2 = 67%).

Sensitivity Analysis

Study Quality

Sensitivity analysis of only trials with higher-quality scoring (a score of at least five on the Delphi list) revealed a moderate non-statistically significant antidepressant effect for exercise compared to control groups (SMD [random effects] -0.41, 95% CI -0.86, 0.05, p= .08). The I^2 heterogeneity showed a large reduction in this analysis (I^2 = 49%) marginally below the acceptable border of ≤50%. Details of this analysis are illustrated in Figure 3.

Removal of Outlier

The removal of one trial[20] that was considered an outlier with regard to effect size resulted in a statistically significant small overall treatment effect for exercise (SMD [random effects] -0.28, 95% CI -0.53, -0.03) with substantially reduced heterogeneity ($I^{2=}$20%). See Figure 4 for this analysis.

Clinical Sample

The inclusion of only trials with clinical samples revealed a moderate statistically significant antidepressant effect for exercise compared to control groups on depressive scores (SMD [random effects] - 0.43, 95% CI -0.84, -0.02, p= .04). The I^2 heterogeneity in this analysis was 44%, indicating an acceptable level (<50%). Details can be found in Figure 5.

General Population Sample

The inclusion of only trials with samples from the general population showed a non-significant moderate effect for exercise compared to control groups (SMD [random effects] -0.52, 95% CI -1.30, 0.26, p= .19). The heterogeneity in this analysis was high (I^2 = 83%). See Figure 6 for this analysis.

INSERT FIGURES 2-6 HERE

Narrative Description of Trials' Reported Findings

Nine trials conducted a between-group comparison of post-intervention depressive symptoms. Of these trials, seven reported no statistically significant differences on depressive symptoms,[16-18,21,22,25,26,29] and two reported statistically significant differences in favour of exercise.[20,24] One trial with sub-analysis of only those participants with elevated baseline depressive symptoms reported significant between-group differences in favour of the exercise condition.[21]

Discussion

This meta-analysis has found a statistically significant treatment effect on depressive symptoms of approximately half a standard deviation for exercise over comparisons in adolescents (SMD=-0.48). Nevertheless, significant heterogeneity was recorded (I^2=67%). After conducting sensitivity analysis to remove an outlier[20], heterogeneity was substantially decreased to a low level (I^2=20%). This is an indication of a considerable consistency for the antidepressant effect of exercise. The finding remained statistically significant; however, the magnitude of the effect was reduced to small (SMD=0.28). Importantly, the removed trial targeted young people residing in a youthful offenders' rehabilitation setting. Therefore, there are likely two critical differences differentiating this trial from the remaining trials in the review and thus may offer an explanation for the outlying status of this trial. First, youthful offenders living in such settings are generally reported to have numerous mental health needs,[30] and second, participation in this trial intervention was compulsory.

When sensitivity analysis of the four studies with clinical samples was undertaken, a statistically significant, moderate effect for exercise was found (SMD=-0.43). Moreover, the lower levels of heterogeneity (I^2= 44%) indicated, as in the above sensitivity analysis, consistency of outcome across all clinical studies. These findings are congruent with recent reviews of RCTs of exercise for adults with depression,[31,32] and recent epidemiological data reporting that the odds of experiencing symptoms of depression such as sadness and suicidal ideation are lowered for adolescents who exercise more frequently than their peers.[33] Further optimism towards the efficacy of exercise in adolescent depression is derived from the methodological quality of the trials with clinical populations, as three of the four studies showed higher-quality scoring. Conversely, in trials without clinical samples, a non-statistically significant moderate effect was recorded (SMD=-0.52), and the methodological quality showed a wider range towards lower scoring. Thus, exercise may be more beneficial to clinical populations over those in the general population. This finding is in contrast to a previous review of exercise for depression in children and young people that found no significant differences between exercise and control conditions for clinical populations.[5]

In the sensitivity analysis of the five trials with higher-quality scores that included three out of the four clinical trials, a moderate treatment effect for exercise was recorded (SMD=-0.41). Heterogeneity in this analysis was I^2=49%, marginally within the acceptable range of <50%. This is an indication of consistency for the antidepressant effect of exercise in trials with higher-quality scoring, although this effect was marginally non-statistically significant. Sensitivity analyses of higher-quality trials not retaining statistical significance are not unusual and have been reported in a recent systematic review of exercise for adults with depression.[31] In contrast to previous reviews, however, in our review, the effect size remained relatively unchanged, and the small number of participants likely influenced the statistical significance. Despite these positive considerations, it should be underlined that most of our

reviewed trials did not adequately describe allocation concealment procedures or blind outcome assessors. Moreover, all but three trials lacked ITT analysis. A further consideration is that all trials employed self-reported outcome measures. Notwithstanding the various potential shortcomings involved (e.g., demand characteristics, social desirability), self-reported outcome measures are widely employed as they are cost- and time-effective, and avoid problems seen in individualised- or group-based interviews conducted by clinicians, researchers, or teachers (e.g. self-consciousness, rapport, or modelling).

Frequency and intensity appear to vary substantially between trials, a finding also reported recently in a systematic review of exercise programme variables for adults with depression.[34] However, all included trials incorporated a group-based supervised exercise in an aerobic form, and the majority of interventions were delivered three times per week, over an average duration of 11 weeks. However, no overall firm recommendations can be presented regarding the optimum intensity and modality of exercise for depression in adolescents. This is primarily due to the limited number of trials reporting such information. Those trials that provided relevant information were exclusively the clinical trials. Specifically, within the four trials that included a clinical sample, three interventions prescribed exercise intensity at varying levels including vigorous, moderate, or low intensity, and one allowed for self-selected intensity that corresponded to low intensity. Although no firm recommendations can be made regarding exercise intensity for clinical populations due to the limited number of trials, it seems that all levels of exercise intensity appear to be effective for symptom reduction in both severe and moderate depression. However, low- and moderate-intensity exercise was employed more frequently. Trials targeting adolescents from the general population in school settings did not report the level of prescribed intensity; as such, it is anticipated that these intervention participants exercised within the typical range of vigorous or moderate intensities seen in school settings in the United States.[35]

Future research should be directed towards the effectiveness of exercise for adolescents with depression that is defined either by diagnosis or cut-off values from validated, clinically relevant outcome measures. Such clinical trials should also explore the comparative effects of low, moderate, and self-selected intensity exercise; and the optimum modality of exercise, with a specific focus on the comparative effect and acceptability of self-selected vs prescribed exercise intensity (and modality), as the former intensity shows better affective responses compared to the latter in both adults and adolescents.[36,37]

This review has updated the literature on the effect of exercise for depression in adolescents aged 13 to 17 years old and is the first to generate synthesised evidence based on trials that have not been synthesised previously. Based on our findings, physical exercise appears to be a promising antidepressant strategy for adolescents aged 13-17 years old. However, few trials employed adequate allocation concealment, blinding of outcome assessors, and ITT analysis. In trials with exclusively clinical samples, in contrast to trials with participants from the general population, a statistically significant moderate antidepressant effect was recorded. The magnitude of this effect combined with the low levels of heterogeneity and the higher methodological quality scoring across these clinical trials allow for more optimism towards the potential use of exercise as a treatment modality for adolescents with depression. However, larger, high-quality trials in clinical practice with adequate description of the exercise interventions in terms of exercise intensity and specific modality are essential before firm conclusions can be drawn. It appears that such trials should ensure that the exercise interventions are group-based and supervised, include an aerobic component, employ moderate- or low-intensity exercise, and be conducted three times per week for a period of approximately seven weeks.

Clinical Guidance

-Notwithstanding the limited clinical research available, exercise appears to be an effective strategy in treating adolescents with elevated levels of depression.

-Group-based and supervised light- or moderate-intensity exercise activities three times a week for a period of between 6 to 12 weeks could bring about an improvement in depression.

-Exercise seems to be equally effective for both moderate and severe depression in both inpatient and outpatient settings.

-More trials with better methodological quality are needed to provide firmer clinical recommendations towards the dose-response relationship.

References

1. Collishaw S, Maughan B, Natarajan L, Pickles A. Trends in adolescent emotional problems in England: a comparison of two national cohorts twenty years apart. *J Child Psychol Psychiatry*. 2010;51:885-894.
2. Wesselhoeft R, Sorensen MJ, Heiervang ER, Bilenberg N. Subthreshold depression in children and adolescents - a systematic review. *J Affect Disord*. 2013;151:7-22.
3. Lewinsohn PM, Hops H, Roberts RE, Seeley JR, Andrews JA. Adolescent psychopathology: I. Prevalence and incidence of depression and other DSM-III-R disorders in high school students. *J Abnorm Psychol*. 1993;102:133-144.
4. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985;100:126-131.
5. Larun L, Nordheim LV, Ekeland E, Hagen KB, Heian F. Exercise in prevention and treatment of anxiety and depression among children and young people. *Cochrane Database Syst Rev*. 2006;(3):CD004691.

6. Brown HE, Pearson N, Braithwaite RE, Brown WJ, Biddle SJ. Physical activity interventions and depression in children and adolescents: a systematic review and meta-analysis. *Sports Medicine*. 2013;43(3):195-206.

7. Lewinsohn PM, Clarke GN, Seeley JR, Rohde P. Major depression in community adolescents: age at onset, episode duration, and time to recurrence. *J Am Acad Child Adolesc Psychiatry*. 1994;33:809-818.

8. Waaktaar T, Borge AI, Fundingsrud HP, Christie HJ, Torgersen S. The role of stressful life events in the development of depressive symptoms in adolescence--a longitudinal community study. *J Adolesc*. 2004;27:153-163.

9. Green H, McGinnity A, Meltzer H, Ford T, Goodman R. *Mental health of children and young people in Great Britain*. London: ONS; 2005.

10. Altman DG, Bland JM. Detecting skewness from summary information. *British Medical Journal*. 1996;313(7066):1200.

11. Higgins JT, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2. 2011; www.cochrane-handbook.org. Accessed 10th May 2014.

12. Verhagen AP, de Vet HCW, de Bie RA, et al. The Delphi List: A Criteria List for Quality Assessment of Randomized Clinical Trials for Conducting Systematic Reviews Developed by Delphi Consensus. *J Clin Epidemiol*. 1998;51:1235-1241.

13. Borenstein M. *Introduction to meta-analysis*. Chichester: Wiley; 2009.

14. Cochrane Effective Practice and Organisation of Care Group. EPOC-specific resources for review authors. http://epoc.cochrane.org/epoc-specific-resources-review-authors. Accessed November 20, 2015.

15. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med*. 2009;151:264-9, W64.

16. Brown SW, Welsh M, Labbe EE, Vitulli WF, Kulkarni P. Aerobic exercise in the psychological treatment of adolescents. *Percept Mot Skills*. 1992;74:555-560.

17. Hughes CW, Barnes S, Barnes C, Defina LF, Nakonezny P, Emslie GJ. Depressed Adolescents Treated with Exercise (DATE): A pilot randomized controlled trial to test feasibility and establish preliminary effect sizes. *Ment Health Phys Act*. 2013;6:119-131.

18. Khalsa SBS, Hickey-Schultz L, Cohen D, Steiner N, Cope S. Evaluation of the mental health benefits of yoga in a secondary school: A preliminary randomized controlled trial. *The Journal of Behavioral Health Services & Research*. 2012;39:80-90.

19. Melnyk BM, Jacobson D, Kelly S, O'Haver J, Small L, Mays MZ. Improving the Mental Health, Healthy Lifestyle Choices, and Physical Health of Hispanic Adolescents: A Randomized Controlled Pilot Study. *J Sch Health*. 2009;79:575-584.

20. Hilyer JC, Wilson DG. Physical fitness training and counseling as treatment for youthful offenders. *Journal of Counselling Psychology*. 1982;29:292-303.

21. Melnyk BM, Jacobson D, Kelly S, et al. Promoting healthy lifestyles in high school adolescents: A randomized controlled trial. *Am J Prev Med*. 2013;45:407-415.

22. Kanner KD. High versus low-intensity exercise as part of an inpatient treatment program for childhood and adolescent depression. *Diss Abstr Int*. 1991;51:4053.

23. Bonhauser M, Fernandez G, Puschel K, et al. Improving physical fitness and emotional well-being in adolescents of low socioeconomic status in Chile: Results of a school-based controlled trial. *Health Promotion International*. 2005;20:113-122.

24. Roshan VD, Pourasghar M, Mohammadian Z. The efficacy of intermittent walking in water on the rate of MHPG sulfate and the severity of depression. *Iran J Psychiatry Behav Sci*. 2011;5:26-31.

25. Jeong YJ, Hong SC, Lee MS, Park MC, Kim YK, Suh CM. Dance movement therapy improves emotional responses and modulates neurohormones in adolescents with mild depression. *Int J Neurosci*. 2005;115:1711-1720.

26. Carter T, Guo B, Turner D, et al. Preferred intensity exercise for adolescents receiving treatment for depression: a pragmatic randomised controlled trial. *BMC Psychiatry*. 2015;15:247.

27. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

28. McHugh ML. Interrater reliability: the kappa statistic. *Biochemia Medica*. 2012;22:276-82.

29. Bonhauser M, Coronado G, Fernandez G, et al. Improving physical fitness and emotional well-being in adolescents of low socioeconomic status in Chile: results of a school-based controlled trial. *Health Promotion International*. 2005;20:113-122.

30. Lennox C. The health needs of young people in prison. *Br Med Bull*. 2014;112:17-25.

31. Cooney GM, Dwan K, Greig CA, et al. Exercise for depression. *Cochrane Database Syst Rev*. 2013;9:Cd004366.

32. Rethorst CD, Wipfli BM, Landers DM. The antidepressive effects of exercise: a meta-analysis of randomized trials. *Sports medicine (Auckland, N.Z.)*. 2009;39:491-511.

33. Sibold J, Edwards E, Murray-Close D, Hudziak JJ. Physical Activity, Sadness, and Suicidality in Bullied US Adolescents. *J Am Acad Child Adolesc Psychiatry*. 2015;54:808-815.

34. Stanton R, Reaburn P. Exercise and the treatment of depression: A review of the exercise program variables. *J Sci Med Sport*. 2014;17:177-182.

35. Kohl H, Cook H. *Educating the Student Body: Taking Physical Activity and Physical Education to School*. Washington, DC: The National Academies Press; 2013.

36. Rose EA, Parfitt G. A quantitative analysis and qualitative explanation of the individual differences in affective responses to prescribed and self-selected exercise intensities. *Journal of Sport and Exercise Psychology*. 2007;29:281-309.

37. Hamlyn-Williams C, Freeman P, Parfitt G. Acute affective responses to prescribed and self-selected exercise sessions in adolescent girls: an observational study. *BMC Sports Science, Medicine and Rehabilitation*. 2014;6:35.

Figure 1: Preferred reporting items for systematic reviews and meta-analyses (PRISMA) flow diagram. Note: RCT = randomized controlled trial.

Figure 2: Meta-analysis of depressive symptom score. Note: Std. = standard.

Figure 3: Sensitivity analysis of higher quality trials (≥5 on Delphi list). Note: Std. = standard.

Figure 4: Sensitivity analysis of outlier removal. Note: Std. = standard.

Figure 5: Sensitivity analysis of clinical sample. Note: Std. = standard.

Figure 6: Sensitivity analysis of general population sample. Note: Std. = standard.

Table 1: Search Terms by Category (EMBASE)

Population	MeSH search: exp adolescent/ or exp child/students/ pediatrics/ Title search: boy$1 or girl$1 or kid$1 or school$ or preschool$ or juvenil$ or under?age$ or teen$ or minor$ or pubescen$ or young people or young person$ or youth$ or student$).m_titl. child$ or adolescen$ or pediatric$ or paediatric$).m_titl
Outcome	MeSH search: Mental health/mental disorders/ or adjustment disorders/ or affective disorders/ or child mental disorders/ or dissociative disorders/ or eating disorders/ or neurotic disorders/ or sleep disorders/ Title search: sadness or low mood or melanchol* or depress* or dysphor* or dysthymi*
Intervention type	MeSH search: exp exercise/ or leisure activities/ or physical fitness/ or exp sports/ exp exercise therapy/ muscle stretching exercises/ dance therapy/ running/ or walking/ Title search: exercis* or physical activit* or sport* or athletic*).m_titl.
Study design limits	Above search limited to: English language and humans and (clinical trial all, or clinical trial or comparative study or controlled clinical trial or pragmatic clinical trial or randomized controlled trial)
Note: MeSH = Medical Subject Heading.	

Table 2: Characteristics of Included Studies

Primary Author	Participants (N; mean age [years]; population; severity of depression)	Intervention	Control/baseline exercise information	Depression measure	Outcome of depressive symptom analysis
Bonhauser (2005)[23]	198; 15.5; High school students in Chile; average (below mild)	Modality – stretching, weight transfer activities (i.e. running) and sports practice Frequency – 90 minutes, three times per week Duration – 40 weeks Intensity – Not reported Attendance – 87% (Mandatory)	Continued regular exercise classes as part of routine education. Once per week. No information on baseline exercise levels.	HADS	No significant between group differences on change scores at post intervention.
Brown (1992)[16]	27; 15.6; Psychiatrically institutionalised patients; primarily for dysthymia and conduct disorder; mild (speculative as poorly reported)	Modality – Running/aerobic exercises alongside their regular exercise classes given as part of routine treatment Frequency – Three times per week Duration - Nine weeks Intensity – Not reported Attendance – Not reported	Continued regular exercise classes as part of routine treatment, frequency unknown. No information on baseline exercise levels.	BDI	No significant differences observed for depressive symptoms between groups at post evaluation or 4-week follow-up.

Carter (2015)[26]	87; 15.4; receiving treatment for depression; severe	Modality – Aerobic and strengthening exercise in form of circuit training Frequency – 45 minutes, two times per week Duration - 6 weeks Intensity – Preferred intensity Attendance – 66%	Continued treatment as usual which did not include exercise, described as sedentary at baseline.	CDI-2	No significant between group difference at post intervention
Hilyer (1982)[20]	60; 17; adolescents in a state industrial youth offenders school; moderate	Modality - fitness training with a fitness counsellor alongside regular physical education classes Frequency – 90 minutes, three times per week Duration - 20 weeks Intensity – Not stated. Intensity – Not reported Attendance – Not reported	Regular physical education programme as part of routine treatment, frequency unknown.	BDI	Significant between group difference in depressive scores at post intervention in favour of the intervention group (p=.03)
Hughes (2013)[17]	26; 17; adolescents meeting *DSM-IV-TR* diagnosis of non-psychotic major depressive disorder; moderate	Modality – supervised (treadmill, stationary bike) and unsupervised (preferred physical activity) Frequency – 30-40 minutes, three times per week Duration – 12 weeks Intensity (vigorous) - minimum of 12 kilocalories/kilogram/week (KKW) energy expenditure by end of intervention Attendance – 77%	Equivalent contact stretching control. No information on baseline exercise levels.	CDRS-R	No significant between group difference at post intervention. Significant differences observed between groups in terms of response (50% reduction in baseline scores) at week 12 (p=.019) and remission (no residual symptoms with CDRS ≤ 28 and CGI-I ≤ 2) (p=.04)
Jeong	40; 16; female students scoring higher on BDI	Modality – Dance Movement Therapy Frequency – 45 minutes, three times	No treatment control.	Measurement of	Significant reduction in depressive symptoms in the

(2005)[25]	and demonstrating depression symptoms over 4-week period; mild	per week Duration - 12 weeks Intensity – Not reported Attendance – Not reported	Participants had not exercised regularly in previous six months.	Psychological Distress (SCL-90-R)	intervention group from baseline to 12 week follow up with a significant main effect for group x time (p<.001).
Kanner (1991)[22]	68; 13.7; patients at a psychiatric treatment centre; moderate/severe	Modality – Dance Movement Therapy Frequency – 60 minutes, three times per week Duration - 8 weeks Intensity – Two separate intervention arms: (a) Low intensity (< 60% of maximum heart) (b) High Intensity (70-85% maximal heart rate) Attendance – Not reported	Equivalent contact, no exercise control. No information regarding baseline exercise levels.	CDI	No significant between group differences for either condition (a) or (b).
Khalsa (2012)[18]	121, 16.8, high school students; not clear	Modality – Yoga class Frequency – 30-40 minutes, three or four times per week Duration -11 weeks Intensity – Not reported Attendance – 73%	Physical education classes as usual, three times per week. No information on baseline exercise levels.	BASC-2 Depression subscale	No significant between group difference on change scores between control and intervention at post intervention (p=.57)
Melnyk (2009)[19]	19; 15.5; Adolescents attending one of two health courses at predominantly Hispanic	Modality – Exercise was part of a 50 minute educational and CBT class and included Frisbee, kickball and walking. Frequency – 15/20 minutes.	Equivalent contact control- Received instruction on health topics, no	BYI-II	No significant within-group difference for intervention or control at post intervention

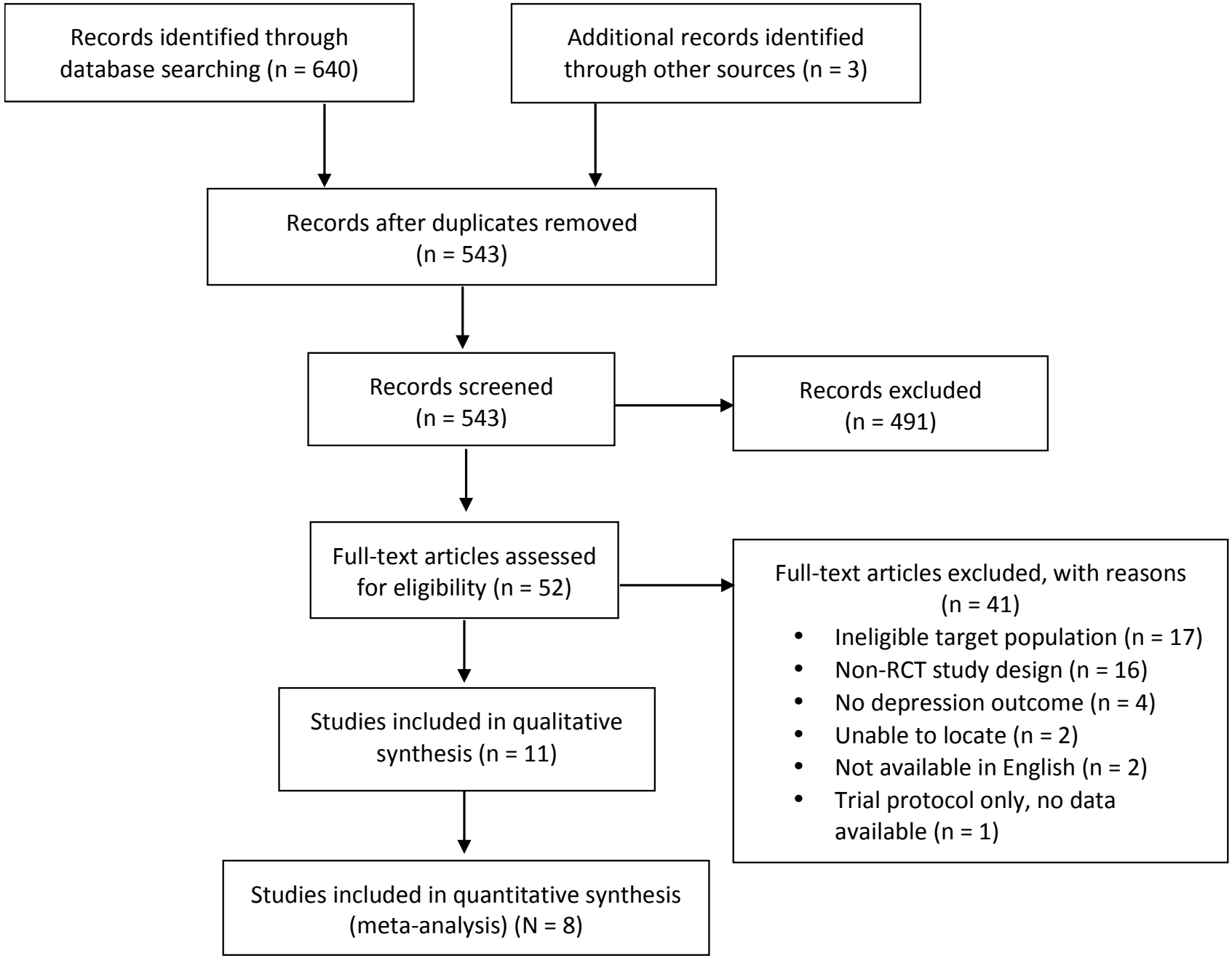
	high school.	Two/Three times a week Duration – Nine weeks Intensity – Not reported Attendance – 92% (Mandatory)	physical activity component. Assumed to engage in normative physical education classes within high school setting		
Melnyk (2013)[21]	779; 14.7; High school students in the south western US; average (below mild)	Modality – Exercise was part of a 50 minute educational and CBT class and included: Frisbee, kickball and walking. Also encouraged to record and increase number of weekly steps by 10% using pedometer. Frequency – 15/20 minutes Duration – 15 weeks, once a week Intensity – Not reported Attendance – Not reported	Equivalent contact time - Received instruction on health topics, no physical activity component. Did have pedometer. Assumed to engage in normative physical education classes within high school setting	BYI-II	No significant differences on depressive symptoms between groups at post intervention (p=.21) Significant between group difference between participants with elevated baseline depression symptoms (p=.02)
Roshan (2011)[24]	24; 16.8; Female students with MDD; severe	Modality - Pool walking with gradual increase in distance throughout the intervention period	No treatment control. No information on	The Hamilton Depression rating scale	Significant difference between groups at post intervention (p<.001)

		Frequency – Three times per week Duration – Six weeks Intensity – 60-70% of maximum heart rate Attendance – Not reported	baseline exercise levels.		

Note: BASC-2 = Behaviour Assessment System for Children, Second Edition; BDI = Beck Depression Inventory; BYI-II = Beck Youth Inventory Second Edition; CDI = Children's Depression Inventory; CDRS-R = Children's Depression Rating Scale – Revised; HADS = Hospital anxiety depression scale; MDD = major depressive disorder.

Table 3: Delphi List

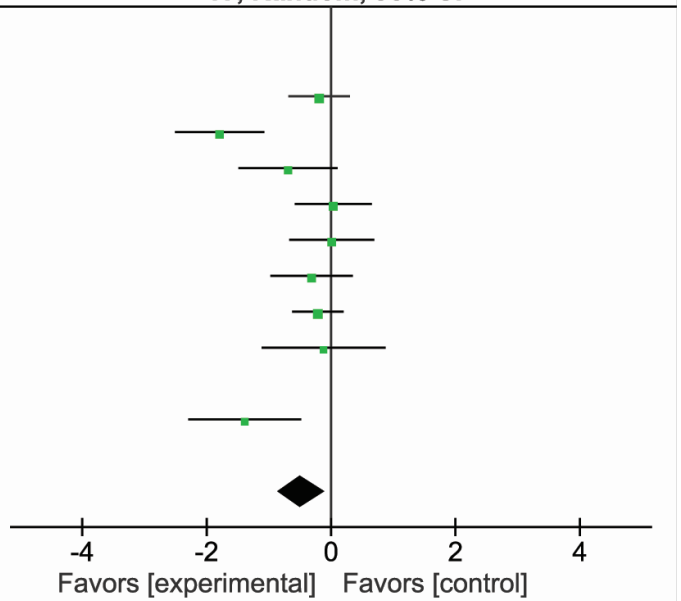
Delphi Criteria	Bonhauser (2002)[23]	Brown (1992)[16]	Carter (2015)[26]	Hilyer (1982)[20]	Hughes (2013)[17]	Jeong (2005)[25]	Kanner (1991)[22]	Khalsa (2012)[18]	Melynk (2009)[19]	Melynk (2013)[21]	Roshan (2011)[24]
Was a method of randomization performed?	1	1	1	1	1	1	1	1	1	1	1
Was the treatment allocation concealed?	0	0	1	0	0	1	0	0	0	0	0
Were the groups similar at baseline regarding the most important prognostic indicators?	1	0	1	1	1	1	1	1	1	0	1
Were the eligibility criteria specified?	1	0	1	0	1	1	1	1	1	1	1
Was the outcome assessor blinded?	0	0	1	0	1	0	0	0	0	0	0
Was the patient blinded?	0	0	0	0	0	0	0	0	1	0	0
Were point estimates and measures of variability presented for the primary outcome measures?	1	0	1	1	1	1	1	1	1	1	1
Did the analysis include an intention-to-treat analysis?	1	0	1	0	0	1	0	0	0	1	1
Total Delphi Score /8	5	1	7	3	5	6	4	4	5	4	5



Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bonhauser (2002)	0	0	0	0	0	0		Not estimable
Brown (1992)	0	0	0	0	0	0		Not estimable
Carter (2015)	23.8	10.7	36	25.7	8.5	28	13.4%	-0.19 [-0.69, 0.30]
Hilyer (1982)	14.43	6.65	23	26.3	6.33	20	10.8%	-1.79 [-2.51, -1.07]
Hughes (2013)	24.1	5.8	14	28.3	5.98	12	9.9%	-0.69 [-1.49, 0.11]
Jeong (2005)	46.4	10.2	20	46.1	5.7	20	11.9%	0.04 [-0.58, 0.66]
Kanner (1991a)	11.9	7.2	17	11.8	11.4	16	11.2%	0.01 [-0.67, 0.69]
Kanner (1991b)	9	5.9	20	11.8	11.4	16	11.4%	-0.31 [-0.97, 0.35]
Khalsa (2012)	45.9	7.31	67	47.5	7.85	34	14.4%	-0.21 [-0.63, 0.20]
Melnyk (2009)	50.64	9.87	11	52.17	15.33	6	8.0%	-0.12 [-1.12, 0.87]
Melnyk (2013)	0	0	0	0	0	0		Not estimable
Roshan (2011)	14.08	5.79	12	25.58	9.72	12	8.8%	-1.39 [-2.30, -0.48]
Total (95% CI)			**220**			**164**	**100.0%**	**-0.48 [-0.87, -0.10]**

Heterogeneity: Tau² = 0.22; Chi² = 24.58, df = 8 (P = 0.002); I² = 67%

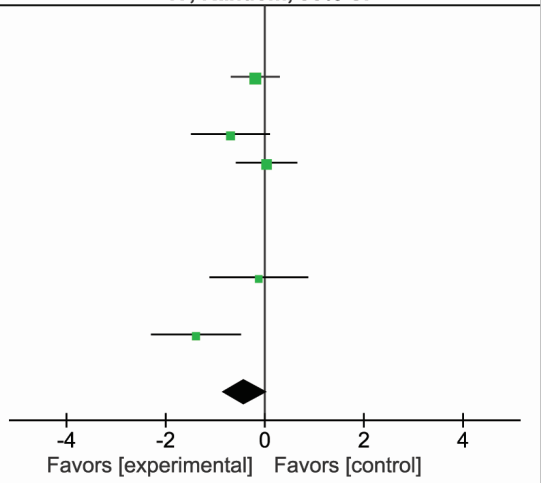
Test for overall effect: Z = 2.45 (P = 0.01)



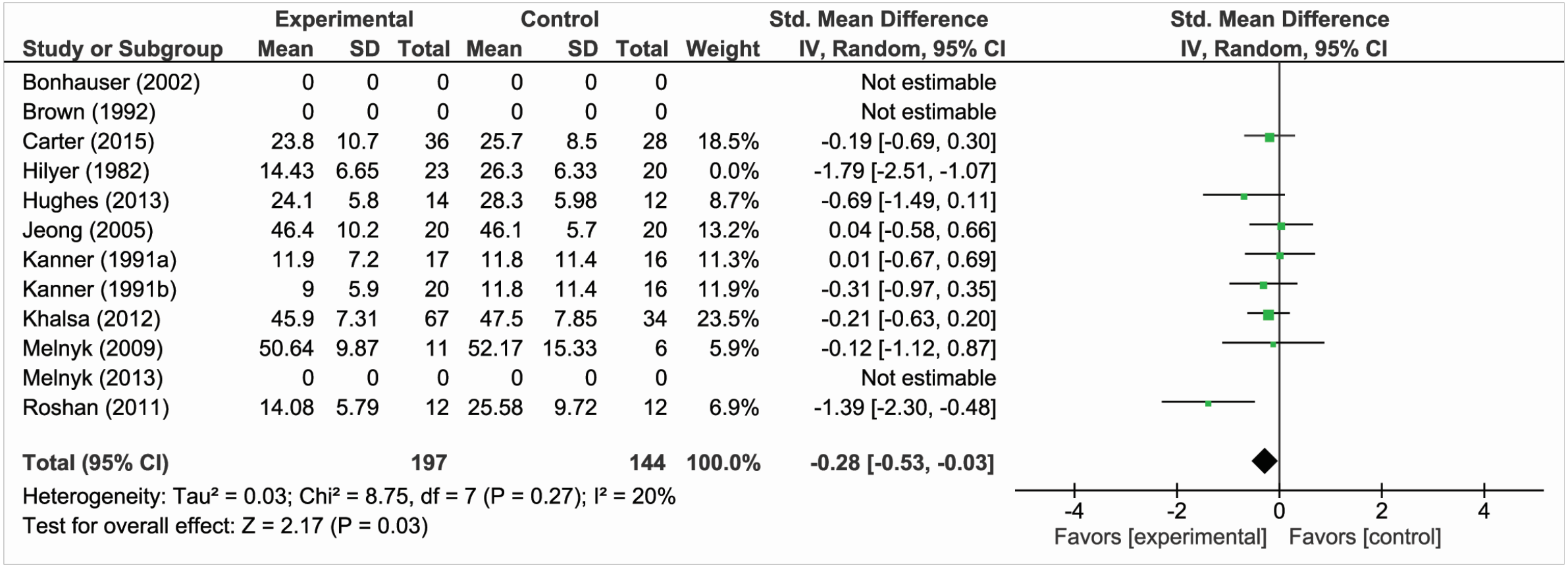
Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bonhauser (2002)	0	0	0	0	0	0		Not estimable
Brown (1992)	0	0	0	0	0	0		Not estimable
Carter (2015)	23.8	10.7	36	25.7	8.5	28	28.2%	-0.19 [-0.69, 0.30]
Hilyer (1982)	14.43	6.65	23	26.3	6.33	20	0.0%	-1.79 [-2.51, -1.07]
Hughes (2013)	24.1	5.8	14	28.3	5.98	12	18.4%	-0.69 [-1.49, 0.11]
Jeong (2005)	46.4	10.2	20	46.1	5.7	20	23.7%	0.04 [-0.58, 0.66]
Kanner (1991a)	11.9	7.2	17	11.8	11.4	16	0.0%	0.01 [-0.67, 0.69]
Kanner (1991b)	9	5.9	20	11.8	11.4	16	0.0%	-0.31 [-0.97, 0.35]
Khalsa (2012)	45.9	7.31	67	47.5	7.85	34	0.0%	-0.21 [-0.63, 0.20]
Melnyk (2009)	50.64	9.87	11	52.17	15.33	6	14.0%	-0.12 [-1.12, 0.87]
Melnyk (2013)	0	0	0	0	0	0		Not estimable
Roshan (2011)	14.08	5.79	12	25.58	9.72	12	15.8%	-1.39 [-2.30, -0.48]
Total (95% CI)			**93**			**78**	**100.0%**	**-0.41 [-0.86, 0.05]**

Heterogeneity: Tau² = 0.13; Chi² = 7.80, df = 4 (P = 0.10); I² = 49%

Test for overall effect: Z = 1.76 (P = 0.08)



Figure

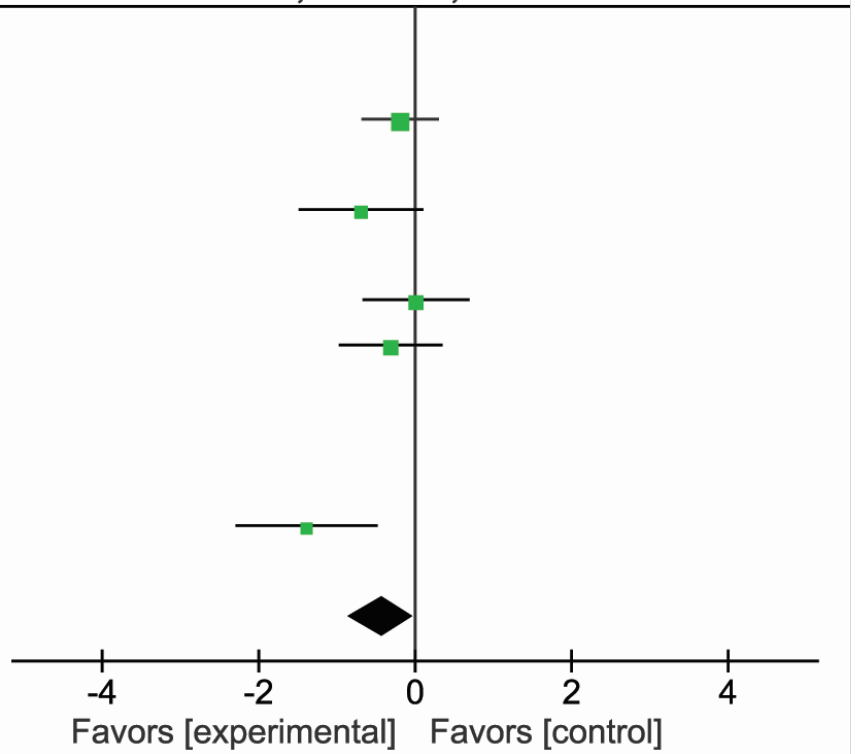

Experimental
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
Bonhauser (2002) 0 0 0 0 0 0 Not estimable
Brown (1992) 0 0 0 0 0 0 Not estimable
Carter (2015) 23.8 10.7 36 25.7 8.5 28 18.5% -0.19 [-0.69, 0.30]
Hilyer (1982) 14.43 6.65 23 26.3 6.33 20 0.0% -1.79 [-2.51, -1.07]
Hughes (2013) 24.1 5.8 14 28.3 5.98 12 8.7% -0.69 [-1.49, 0.11]
Jeong (2005) 46.4 10.2 20 46.1 5.7 20 13.2% 0.04 [-0.58, 0.66]
Kanner (1991a) 11.9 7.2 17 11.8 11.4 16 11.3% 0.01 [-0.67, 0.69]
Kanner (1991b) 9 5.9 20 11.8 11.4 16 11.9% -0.31 [-0.97, 0.35]
Khalsa (2012) 45.9 7.31 67 47.5 7.85 34 23.5% -0.21 [-0.63, 0.20]
Melnyk (2009) 50.64 9.87 11 52.17 15.33 6 5.9% -0.12 [-1.12, 0.87]
Melnyk (2013) 0 0 0 0 0 0 Not estimable
Roshan (2011) 14.08 5.79 12 25.58 9.72 12 6.9% -1.39 [-2.30, -0.48]
Total (95% CI) 197 144 100.0% -0.28 [-0.53, -0.03]
Heterogeneity: Tau² = 0.03; Chi² = 8.75, df = 7 (P = 0.27); I² = 20%
Test for overall effect: Z = 2.17 (P = 0.03)
-4
-2
0
2
4
Favors [experimental]
Favors [control]

Figure

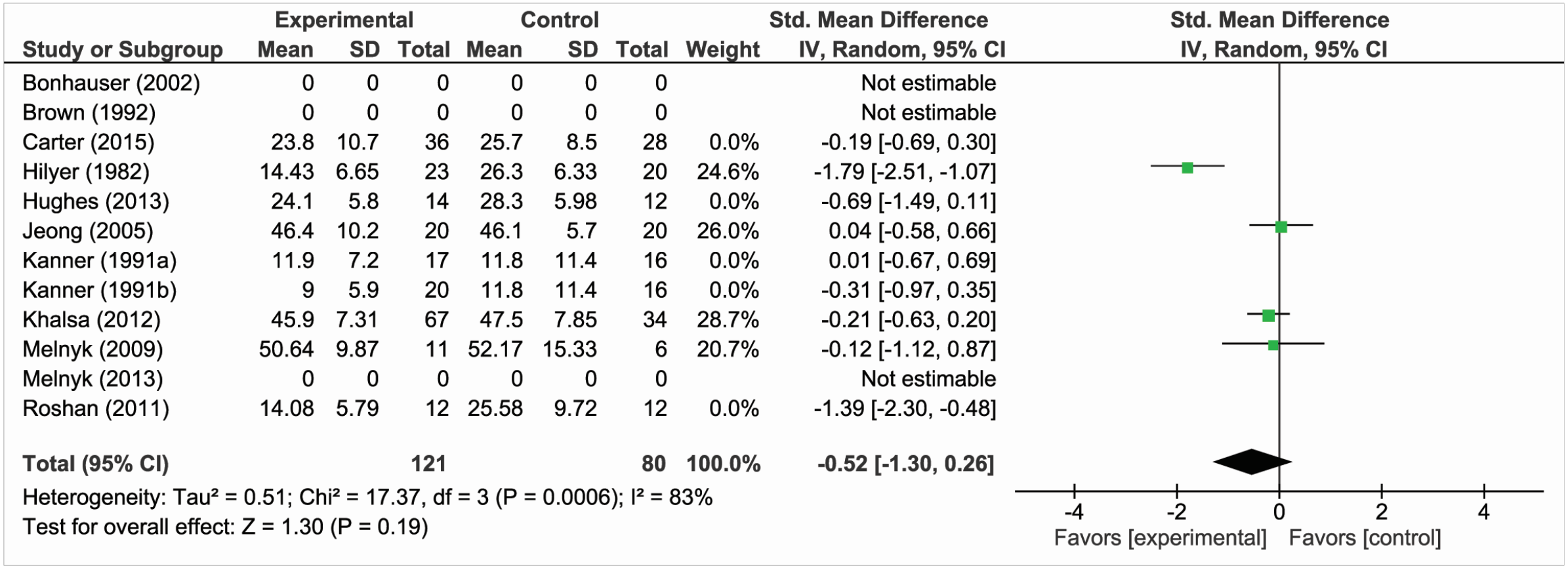
Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Bonhauser (2002)	0	0	0	0	0	0		Not estimable
Brown (1992)	0	0	0	0	0	0		Not estimable
Carter (2015)	23.8	10.7	36	25.7	8.5	28	27.6%	-0.19 [-0.69, 0.30]
Hilyer (1982)	14.43	6.65	23	26.3	6.33	20	0.0%	-1.79 [-2.51, -1.07]
Hughes (2013)	24.1	5.8	14	28.3	5.98	12	16.9%	-0.69 [-1.49, 0.11]
Jeong (2005)	46.4	10.2	20	46.1	5.7	20	0.0%	0.04 [-0.58, 0.66]
Kanner (1991a)	11.9	7.2	17	11.8	11.4	16	20.3%	0.01 [-0.67, 0.69]
Kanner (1991b)	9	5.9	20	11.8	11.4	16	21.0%	-0.31 [-0.97, 0.35]
Khalsa (2012)	45.9	7.31	67	47.5	7.85	34	0.0%	-0.21 [-0.63, 0.20]
Melnyk (2009)	50.64	9.87	11	52.17	15.33	6	0.0%	-0.12 [-1.12, 0.87]
Melnyk (2013)	0	0	0	0	0	0		Not estimable
Roshan (2011)	14.08	5.79	12	25.58	9.72	12	14.2%	-1.39 [-2.30, -0.48]
Total (95% CI)			**99**			**84**	**100.0%**	**-0.43 [-0.84, -0.02]**

Heterogeneity: Tau² = 0.10; Chi² = 7.17, df = 4 (P = 0.13); I² = 44%

Test for overall effect: Z = 2.05 (P = 0.04)



Figure


Experimental
Control
Std. Mean Difference
Std. Mean Difference
Study or Subgroup Mean SD Total Mean SD Total Weight IV, Random, 95% CI
IV, Random, 95% CI
Bonhauser (2002) 0 0 0 0 0 0 Not estimable
Brown (1992) 0 0 0 0 0 0 Not estimable
Carter (2015) 23.8 10.7 36 25.7 8.5 28 0.0% -0.19 [-0.69, 0.30]
Hilyer (1982) 14.43 6.65 23 26.3 6.33 20 24.6% -1.79 [-2.51, -1.07]
Hughes (2013) 24.1 5.8 14 28.3 5.98 12 0.0% -0.69 [-1.49, 0.11]
Jeong (2005) 46.4 10.2 20 46.1 5.7 20 26.0% 0.04 [-0.58, 0.66]
Kanner (1991a) 11.9 7.2 17 11.8 11.4 16 0.0% 0.01 [-0.67, 0.69]
Kanner (1991b) 9 5.9 20 11.8 11.4 16 0.0% -0.31 [-0.97, 0.35]
Khalsa (2012) 45.9 7.31 67 47.5 7.85 34 28.7% -0.21 [-0.63, 0.20]
Melnyk (2009) 50.64 9.87 11 52.17 15.33 6 20.7% -0.12 [-1.12, 0.87]
Melnyk (2013) 0 0 0 0 0 0 Not estimable
Roshan (2011) 14.08 5.79 12 25.58 9.72 12 0.0% -1.39 [-2.30, -0.48]
Total (95% CI) 121 80 100.0% -0.52 [-1.30, 0.26]
Heterogeneity: Tau² = 0.51; Chi² = 17.37, df = 3 (P = 0.0006); I² = 83%
Test for overall effect: Z = 1.30 (P = 0.19)
-4
-2
0
2
4
Favors [experimental]
Favors [control]