**Challenges establishing the efficacy of exercise as an antidepressant treatment: a systematic review and meta-analysis of control group responses in exercise randomised controlled trials**

**Accepted by Sports Medicine**

Brendon Stubbsa,b, Davy Vancampfortc,d, Simon Rosenbaume, Philip B Warde, Justin Richardsf, Michael Ussherg, Felipe B Schuchh,i

**Running title:** Control group responses in exercise and depression studies

aPhysiotherapy Department, South London and Maudsley NHS Foundation Trust, Denmark Hill, London SE5 8AZ, United Kingdom

bHealth Service and Population Research Department, Institute of Psychiatry, King's College London, De Crespigny Park, London, Box SE5 8AF, United Kingdom

cKU Leuven – University of Leuven Department of Rehabilitation Sciences, Leuven, Belgium

dKU Leuven – University of Leuven, Z.org Leuven, campus Kortenberg, Kortenberg, Belgium

e School of Psychiatry, University of New South Wales, Sydney, Australia

fSchool of Public Health, Charles Perkins Centre, University of Sydney, Sydney, Australia

g Population Health Research Institute, St. George’s University of London, London, United Kingdom

h Hospital de Clinicas de Porto Alegre, Porto Alegre, Brazil

iProgramade Pós Graduação em Ciências Médicas: Psiquiatria, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

**Corresponding author**: Brendon Stubbs, Physiotherapy Department, South London and Maudsley NHS Foundation Trust, Denmark Hill, London, United Kingdom. Tel: 0044 208 3003100, fax 00442032282702 email:[brendonstubbs@hotmail.com](mailto:brendonstubbs@hotmail.com)

**Abstract**

**Background**

Whilst previous meta-analyses have demonstrated that control group responses (CGRs) can negatively influence antidepressant efficacy, no such meta-analysis exists in exercise randomised control trials (RCTs).

**Objective**

Conduct a systematic review and meta-analysis investigating CGRs and predictors in control groups of exercise RCTs, among adults with depression.

**Methods**

Three authors acquired RCTs from a previous Cochrane review (2013) and conducted updated searches of major databases from 01/2013 to 08/2015. We included exercise RCTs that: (1) involved adults with major depressive disorder (MDD) or depressive symptoms; (2) measured depressive symptoms pre and post-intervention using a validated measure (e.g., Hamilton Depression Scale (HAM-D); (3) included a non-active control group. A random effects meta-analysis calculating the standardised mean difference (SMD) together with 95% confidence intervals (CIs) was employed to determine CGR.

**Results**

Across 41 studies, 1,122 adults with depression were included (mean (SD) age 50 (18) years, 63% female). A large CGR of improved depressive symptoms was evident across all studies (SMD -0.920, 95% CI=-1.11 to -0.729). CGR were elevated across all subgroup analyses, including high quality studies (N=11, SMD=-1.430, 95% CI -1.771 to -1.090) and MDD participants (N=23 SMD=-1.161, 95%CI=-1.410 to -0.912). The CGR equated to an improvement of -7.5 points on the HAM-D (95% CI -10.30 to -4.89). In MDD participants, increasing age moderated a smaller CGR, while the percentage of drop-outs, baseline depressive symptoms and a longer control group duration moderated a larger CGR (i.e. improvement) (p<0.05).

**Conclusion**

In order to demonstrate effectiveness, exercise has to overcome a powerful CGR of approximately double that reported for antidepressant RCTS.

**Key Points**

* Control groups used in exercise and depression randomized clinical trials experience large improvements in depressive symptoms.
* The improvement in control groups equates to approximately double that reported in control groups of antidepressant studies.
* Exercise has to ‘beat’ a powerful control group arm in order to demonstrate its effectiveness against depression.

1. **Introduction**

Major depressive disorder (MDD) is a serious public health concern, ranking second in the top ten causes of years lived with disability in all countries in the recent Lancet review [1]. Moreover, MDD and sub-threshold depressive symptoms are pervasive, affecting people of all ages, sexes and across a broad spectrum of chronic conditions [2]. Treatment of depression is multifaceted, often including both pharmacological therapy (e.g., antidepressants) and psychotherapies (e.g., cognitive behavioural therapy). However, remission rates following first treatment are often poor with approximately 37%, requiring further intervention [3]. In recent years, the notion of ‘exercise is medicine’ for people with depression has received considerable attention [4-6], and treatment guidelines proposed exercise as a therapeutic strategy for managing depressive symptoms [7]. The benefits of exercise for people with depression extend beyond alleviating depressive symptoms [4]. For instance, people with depression are at increased risk of diabetes [8] and premature mortality from cardiovascular disease [9] and, in the general population, exercise is broadly as effective as pharmacological interventions in preventing cardiovascular disease and related premature mortality [10]. Both depressive symptoms and cardio-metabolic disease are associated with lower levels of physical activity participation compounding the risk of high levels of sedentary behaviour among people with depression [11].

Evidence regarding the antidepressant effect of exercise for people with depression is mixed. Whilst some RCTs have reported a positive impact [6, 12] others have found no significant treatment effect [13, 14] and it is important to understand why these outcomes are inconsistent [15, 16].

Randomised controlled trials (RCTs) using the comparison of active interventions with control groups that receive either placebo, treatment as usual (TAU), or a non-active intervention, have been the basis of evidence-based medicine for many years. In the case of medication, control groups typically receive an inert pill with no active ingredient as the placebo [15] and, in order to demonstrate its effectiveness, the active medication has to show evidence of a treatment effect beyond that seen in the placebo response. The placebo response is defined as the change in illness symptoms occurring during a clinical trial in patients randomized to receive placebo [17, 18]. There are concerns about increasing evidence of pronounced placebo responses in studies of antidepressant medication, making it more challenging for novel medications to demonstrate effectiveness [17, 18]. A large placebo response has also been reported in non-pharmacological interventions for depression, with a recent meta-analysis demonstrating an effect size of 0.82 (95% CI 0.63 to 1) for control group responses in trials of repetitive transcranial magnetic stimulation [19].

Recently, a meta-analysis [20] quantified the placebo response for psychological outcomes in exercise studies in the general population, finding a mean effect size of 0.20 (95% CI 0.02 to 0.41). In exercise trials among people with depression, control group participants do not receive a conventional placebo. For instance, many people in control groups are in receipt of antidepressant medication and/or receiving psychotherapy. Thus, TAU is frequently a proven and effective ‘intervention’ for depressive symptoms. Therefore, while RCTs of antidepressant medication typically contend with placebo responses of inert pills, exercise interventions, when added to TAU, have to demonstrate greater symptom reductions than potent control interventions, and failure to do so may be interpreted as a lack of efficacy. Thus, in exercise studies among patients with depression, it is important to quantify the control group response, and to consider the ‘actual’ effect of exercise interventions. This information is critical in designing appropriate control interventions, in guiding clinical decision-making, and in facilitating policy makers in funding cost-effective interventions.

A recent Cochrane review [21] found that exercise improved symptoms of depression (standardised mean difference (SMD) -0.62, 95% CI -0.81 to -0.42). However the results were criticised for questionable restrictions to high quality trials resulting in a shrinkage of the effect size [22]. Moreover, the extent to which control group responses reduced estimates of the effectiveness of exercise interventions was not considered. Therefore, we conducted the first systematic review and meta-analysis to quantify the control group response in studies of exercise interventions for people with depression. Additionally, we used sensitivity and meta-regression analyses to identify factors that may contribute to the control group response.

1. **Methods**

This systematic review adhered to the MOOSE guidelines [23] and PRISMA statement [24], following a predetermined published protocol (PROSPERO registration CRD42015025333).

*2.1 Inclusion criteria*

Included in this meta-analysis were studies that:

(1) Included adult participants with a primary diagnosis of MDD according to established criteria (e.g., DSM-IV, [25] or ICD-10, [26]) or those with increased depressive symptoms determined by a validated screening measure (e.g., Hamilton Rating Scale for Depression, HMD-R), [27], Beck Depression Inventory (BDI or BDI-II [28]). We also included studies meeting our criteria among people with MDD who had other comorbid diagnosis such as bipolar disorder and dysthymia which was decided among three reviewers (BS, FS, SR).

(2) Measured depressive symptoms pre- and post-intervention using a validated measure (e.g., HAM-D, BDI).

(3) Were RCTs investigating exercise as the active arm of the trial.

(4) Included a control group, defined as either: receiving e.g. TAU defined as either: usual-care, antidepressants, psychotherapies, electroconvulsive therapy, or wait-list control conditions, or in line with the recent Cochrane review [21] also any light intensity activity programs (e.g., stretching, walking).

(5) Were published in a peer-review journal or a published PhD dissertation.

*2.2 Information sources and searches*

Articles were identified in a two-step strategy. First, three authors (BS, FS, SR) reviewed all articles identified (both included and excluded with reasons) by the recent Cochrane review on exercise for depression [21]. Second, three independent reviewers (BS, FS, SR) searched Academic Search Premier, MEDLINE, Psychology and Behavioral Sciences Collection, PsycINFO, SPORTDiscus, CINAHL Plus and PubMed without language restrictions from database January 2013 until July 1st, 2015, using the key words: ((exercis\* OR aerobic\* OR running OR jogging OR walk\* OR hiking OR swim\* OR aquatic\* OR cycling OR bicycl\* OR strength\* and activit\* OR fitness OR train\* OR "physical medicine" OR resistance OR lift\*) AND (depression OR dysthymia)). In addition, reference lists of all eligible articles of recent reviews investigating the effectiveness of exercise versus control were screened to identify potentially eligible articles [21, 29, 30].

*2.3 Study selection*

In the first stage of the search strategy, three authors (BS, FS, SR) determined potentially eligible articles meeting criteria from the Cochrane review [21]. In the second stage, after removal of duplicates, two independent reviewers screened titles and abstracts of all potentially eligible articles. Three authors applied the eligibility criteria, considered the full texts and a final list of included articles was reached through consensus.

2.4 Outcomes:

Our primary outcome of interest was the mean change in depressive symptoms within the control group from baseline to the end of the intervention, defined as the ‘control group response’.

*2.5 Data Extraction*

Two authors (FS, SR) independently extracted data using a data extraction form, including: study design, geographical location, details of control group participants (mean age, % females, presence of clinical comorbidities, such as cardiovascular, metabolic, neurological conditions (yes/ no)), psychopharmacologic treatment, including % on antidepressants and details regarding depression diagnosis. We also extracted data on the publication type (peer reviewed journal article or dissertation) the duration of the control intervention, type of control condition (i.e., wait list/no intervention, placebo pill, antidepressants, psychotherapies, light intensity activity/stretching, usual treatment/routine care/counselling, meditation, light therapy, ECT), and drop-out frequencies in the control group. Finally, we extracted data on the mean and standard deviation (SD) pre and post-test depressive symptom rating scale for the control group (primary outcome). If this was not available, we used the mean change and SD from pre and post-test if reported within the study.

*2.6 Risk of Bias and quality assessment*

Three authors (FS, JR, BS) assessed studies on the presence of high, low risk or unclear risk of bias according to the Cochrane Handbook definition [31]. The risk of bias was assessed in the following items: random sequence generation, allocation concealment, blinding of participants, blinding of those delivering the intervention, blinding of outcome assessors, incomplete data outcome, selective reporting or others. Studies presenting adequate allocation concealment and complete presentation of outcome data (intention-to-treat analysis) and blinding outcome assessors are considered studies with low risk of bias (high quality trials). The criteria selection was based on previous studies [21].

*2.7 Meta-analysis*

Due to the anticipated heterogeneity, we utilised random effects meta-analysis and calculated SMD and 95% confidence intervals (CIs) as the effect size measure (ES). The meta-analysis was conducted in the following steps. First, we calculated the SMD statistic together with 95% CI to establish the control group response in our composite analysis across all studies using Comprehensive Meta-Analysis software (CMA; Version 3, Biostat, Englewood, New Jersey). We subsequently conducted a sensitivity analysis computing the control group response in studies using a wait list or treatment as usual only in order to reduce the potential for any active influence on the control group response that may be present in other control groups employed. Second, we conducted meta-regression analyses with CMA to investigate potential moderators of the control group response. Potential moderators of interest were chosen which have previously been associated with responses in exercise RCTS; these were: mean age, sex, sample size, study publication year, duration of control group and baseline depressive symptomology. We also investigated the influence of frequency of adherence, and drop out in the control groups. We also conducted categorical meta-regression analyses to investigate the potential impact of risk of bias comparing studies with a low risk of bias versus those with a high or unclear risk of bias on key study design factors, including allocation concealment, intention-to-treat analysis, and blinding of assessors. Next, we conducted subgroup analyses to compare control group response according to depression diagnosis (MDD versus depressive symptoms), study setting (inpatient, outpatient, mixed), type of publication (peer review article or thesis), high quality (low risk of bias) versus low quality, and presence of comorbidity in the study participants (yes or no). Within each subgroup analysis we also calculated the control group response using wait list/usual care control groups also. Finally, in line with meta-analysis investigating the placebo response in depression RCTs [17], we investigated mean changes in depressive symptoms with the HAM-D and BDI. Heterogeneity was assessed with the Cochran Q and I2 statistics for each analysis [32]. Publication bias was assessed with a visual inspection of funnel plots and with the Begg-Mazumdar Kendall's tau [33] and Egger bias test [34]. Moreover, for the main composite analysis we conducted a trim and fill adjusted analysis [35] to remove the most extreme small studies from the positive side of the funnel plot, and recalculated the effect size at each iteration, until the funnel plot was symmetric about the (new) effect size. Finally, we calculated the fail safe number of negative studies that would be required to nullify (i.e. make p>0.05) our main effect size [36].

1. **Results**

**3.1 Search results**

In the first stage of our search strategy, 35 RCTs were identified from a previous review [21]. In the second stage, following the removal of duplicates, we identified 819 potentially relevant articles from our searches. At the full text review stage, we reviewed 76 articles (N=35 from stage 1 and 41 from our searches in stage 2) and 30 were excluded with reasons (details summarised in figure 1). There were 46 full texts that met the eligibility criteria. Of these, 41 had complete data to enable inclusion within our meta-analysis, including 32 [12, 21, 37-66] from a previous review [21] and nine new RCTs [14, 67-74].

*Figure 1 here*

*3.2 Characteristics of included trials and participants*

Across the 41 studies, 1,122 adults with depression were included in the control groups of the included studies (mean age 49.8 years (±18), 63% female (range 17-100%)). The mean proportion of antidepressant use across the included studies was 38.1% (range 0-100%). Overall, 23 studies [12, 14, 37, 39, 43-45, 47, 48, 51, 53, 54, 56, 57, 65, 67-70, 72-75] confirmed a diagnosis of MDD, whilst 15 studies included participants with depressive symptoms [38, 40-42, 46, 49, 50, 55, 58, 60-62, 66, 71, 76], and a further three included participants with MDD with a number of people with additional co-morbid diagnosis. Specifically, Singh et al [64] and [63] included participants with MDD and dysthymia, Knubben et al [52] included some participants with bipolar disorder. The majority of included studies involved outpatients with depression (N=36), were published in peer-reviewed journals (N=35), and included people without reported clinical comorbidities (N=35). The most commonly used measures of depressive symptoms were the HAM-D (N=15), BDI (N=10), or BDI II (N=3). Most control groups used a waiting list (N=11) or usual care (N=15). Full details of the included studies, participant details, control group condition and depressive symptom measurement are presented in table 1.

*3.3 Risk of bias*

Overall, eleven studies were judged to be of good methodological quality and at low risk of bias [12, 14, 37, 38, 43, 53, 54, 67, 69, 73, 74] and the remaining 30 were of low quality (high risk of bias). Full details of the risk of bias are presented in Electronic Supplementary Material Table S1.

*Insert table 1 here*

**3.4 Main analysis – control group response across all studies**

Data was pooled from 41 studies and established that the control group experienced a large significant improvement (i.e., reduction) in depressive symptoms (SMD -0.920, 95% CI -1.11 to -0.729, p<0.001, Q=634, p<0.01) (figure 2). The Begg (Tau =-0.32, p=0.003) but not the Egger tests indicated publication bias (intercept =1.5, p=0.2). Therefore, the effect size was recalculated using Duval and Tweedies trim and fill method with five studies being adjusted and a new effect size of -1.09 (95% CI -1.31 to -0.873, p<0.001).

The fail-safe number of additional negative studies (that is, studies reporting no changes in control group improvement in depressive symptoms) required to nullify the significance of the main analysis was high. Indeed, 7,886 studies with negative results were required to nullify the large significant result.

Figure 2 here

**3.5 Sensitivity analysis investigating control group response in wait list and treatment as usual**

In our sensitivity analysis, the control group response in 26 RCTs utilising wait list or usual care was a SMD of -0.642 (95% CI -0.853 to -0.431, Q=332, p<0.01). There was no evidence of publication bias (Egger = 0.35, p=0.7, Begg=-0.26, p=0.055). The fail-safe number was 2,241 studies.

**3.6 Meta-regression of control group response in main analysis**

With borderline significance (p=0.05), higher mean age of control group participants (β=0.0226, 95%CI=-0.0009 to 0.0461, z=1.88) moderated less reduction in depressive symptoms in the control group response and explained some heterogeneity (R²=0.13). The percentage of control group drop-outs also moderated a larger reduction in depressive symptoms (i.e. control group improves) in the control groups (β=-0.023, 95%CI -0.0339 to -0.0134, Z=-4.61, p=0.001, R²=0.16). Baseline depressive symptomology was not related to the control group response (β=-0.0058, 95% CI=-0.0342 to 0.0226, z=-0.4, p=0.688). A diagnosis of MDD moderated a larger improvement in depressive symptoms in the control group compared with studies using depressive symptoms (β=-0.6319, 95% CI -1.2088 to-0.0549, z=-2.15, p=0.03, R²=0.13). A multivariate meta-regression model established that the percentage of drop-outs in the control group (β=-0.0211, 95% CI-0.0388 to-0.0033, z=-2.32, p=0.02) but not a diagnosis of MDD (β=-0.5322, 95% CI -1.2451 to 0.1807, z=-1.46, p=0.14) or mean age (β=0.0175, 95% CI -0.0042 to 0.0392, z=1.58, p=0.11) moderated a larger improvement in the control group response. The multivariate model explained almost half of the heterogeneity in the control group response effect size (R² = 0.43). A summary of all meta regression analyses is presented in table 2.

**3.7 Mean change in depressive symptoms**

Data from 13 studies established that control group participants improved by -7.59 points (95% CI -10.30 to-4.889, p<0.001) on the HAM-D scale. A sensitivity analysis including only studies utilising a wait list or usual care control across seven studies established a mean improvement in depressive symptoms of -4.50 points on the HAM-D (95% CI -8.18 to -0.864, p=0.01). The mean control group improvement in depressive symptoms across 10 studies using the BDI was -4.862 (95% CI -8.374 to -1.350, p=0.007), and -4.167 points (95% CI -8.14 to -0.854, p=0.01) in nine studies using a wait list or usual care.

*3.8 Sensitivity analyses*

*3.8.1 Control group response in MDD and depressive symptoms*

Overall, the pooled control groups response across 23 studies in participants with MDD was -1.161 (95% CI-1.41 to -0.912, Q=354, p<0.01, Begg=-0.26, p=0.08, Egger=-2.19, p=0.13) larger than in studies considering depressive symptoms (SMD =-0.481 95% CI -0.738 to -0.224, N=15, Q=120, p<0.01, Begg=-0.33, p=0.08, Egger=-4.31, p=0.05) (See figure 3). The fail safe number required to nullify the result was 5481.

*Figure 3 here*

The control group response remained significant in studies using wait list/usual care in MDD (SMD -0.796, 95% CI -1.08 to -0.503, p<0.001, N=13) and in those studies investigating depressive symptoms (SMD -0.283, 95% CI -0.505 to -0.062, p=0.01, N=11).

*3.8.2 Meta regression of moderators of control group response in MDD*

Mean age (β=0.0498, 95% CI 0.0108 to 0.0888, z=2.51, p=0.01, R² = 0.34), the percentage of drop outs (β=-0.0197, 95% CI-0.0367 to -0.0027, z=-2.27, p=0.02, R² = 0.21), baseline depressive symptoms (β=-0.0535, 95% CI -0.1067 to-0.0002, z=-1.97, p=0.04, R² = 0.18) and longer control group duration moderated the control group response in MDD studies (β=0.0592, 95% CI 0.0003 to 0.1181, z=1.97, p=0.04, R²=0.17). When all were entered into a multivariate meta-regression model, none were significant moderators but percentage drop-out demonstrated a marginal significance (β=-0.0179, 95% CI -0.0366 to 0.0008, z=-1.88, p=0.06) and the model explained over half of the heterogeneity (R² = 0.59).

*3.8.3 Meta regression of moderators of control group response in depressive symptoms*

No significant moderators were found (see table 2).

**3.9 Study quality**

We pooled the control group response separately for those studies deemed a high/unclear risk of bias separately from low risk of bias and found a larger control group response in high quality studies (SMD -1.430, 95% CI -1.771 to -1.090, p<0.001; Q=167, p<0.01) than lower quality studies (SMD -0.713, 95% CI -0.925 to -0.498, p<0.001; Q=318, p<0.01). Categorical meta-regression analysis found that studies judged high risk of bias on allocation concealment, ITT analyses and assessor blinding were not significant moderators across all studies or in those in MDD alone (p>0.05, the results of these meta regression analysis are available from the first author upon request).

**3.10 Study setting**

A larger control group response was observed in inpatient settings (SMD -1.783, 95% CI -2.405 to -1.161, Q=94, p<0.01) compared with outpatient settings (SMD -0.884, 95% CI -1.048 to -0.639, Q=462, p<0.01, N=36). When restricted to wait list/usual care, the results remained significant in inpatient studies (SMD 1.99, 95% CI -2.78 to -1.26, N=2) and outpatient studies (SMD -0.542, 95% CI -0.748 to -0.337, N=23).

**3.11 Control group response according to type of control group comparison**

All of the control groups demonstrated a significant improvement in depressive symptoms (see table 3).

**3.12 Comorbidity and control group response**

Studies conducted in participants without comorbidities (N=35) demonstrated a larger effect size (SMD -0.978, 95% CI -1.18 to-0.769) than those including people with comorbidities (N=6, SMD -0.607, 95% CI -1.09 to -0.115). The results remained significant for those without (N=20, SMD -0.689, 95% CI -0.94 to -0.438) and with comorbidities (N=5, SMD -0.571, 95% CI -1.05 to -0.08) when restricted to wait list/usual care only.

**3.13 Type of publication**

Control group responses were evident in RCTs in peer review journals (N=34, SMD -0.932, 95% CI -1.13 to -0.725) and theses (N=7, SMD -0.850, 95% CI -1.34 to -0.357). The results were significant in peer review journals when restricted to wait list/ usual care controls (N=21, SMD -0.696, 95% CI -0.920 to -0.472) but not in theses (N=5, SMD -0.37, 95% CI -0.886 to 0.13).

All sensitivity and subgroup analyses are presented in table 3.

*Insert table 2 here*

*Insert table 3 here*

1. **Discussion**

The present review found consistent evidence that control groups in exercise RCTs experience pronounced improvements in depressive symptoms. This large control group improvement was evident across virtually all subgroup analyses, and was highest when we restricted the analyses to high quality studies. Moreover, a large control group response seems evident within participants with confirmed MDD. Of all control group ‘methods’ used, a waiting list (N=11) appears to have the least improvement, although this is still in the medium effect size range. Meta-regression analyses demonstrate that increasing age results in less control group improvement, whilst a higher proportion of drop-outs results in a larger control group improvements. Moreover, higher depressive symptoms also appear to moderate a smaller control group response in people with MDD. Taken together, control groups improve by 7.5 points on the HAM-D scale across all studies, or 4.5 points in treatment as usual/waiting list studies. This demonstrates that control groups in exercise RCTS show an approximate doubling in the improvement in HAM-D scores compared with observations for control groups in meta-analyses of antidepressants RCTs (4 points, [17]), with the National Institute for Health and Care Excellence (NICE) stating an improvement of 3 points is clinically meaningful [77]. Thus, considering the substantial control group response that exercise RCTs must overcome to show a benefit in depressive symptoms and the benefits in absolute terms, critiques concerning the efficacy of exercise for depression (e.g., [13]) may be premature, particularly in light of the plethora of other health benefits of an active lifestyle [10, 78, 79].

There are a number of strengths for this meta-analysis. First, the results are novel and address a critically important area with MDD continuing to be a leading cause of disability worldwide [1]. Second, we conducted a comprehensive systematic review, critical appraisal and robust meta-analysis and meta-regression analyses. Third, our results were consistent across all analyses. We conducted numerous adjustments for publication bias, while the number of studies needed to nullify our results are extremely high, adding support to the robust nature of the current findings. A number of limitations should be considered when interpreting the findings of our meta-analysis. First, we encountered some heterogeneity in most of the analyses. However, we were able to explain large portions of this with our meta-regression analyses. Second, some important participant data was not available across most studies and it was therefore not possible to investigate the influence of these data (e.g. illness duration, psychotropic medication use). In addition, lack of data with a number of RCTs meant they were not eligible for inclusion in our meta-analysis. Third, due to limitations in the dataset we could not consider the long term control group responses in the RCTs in a systematic manner. An important area for future research, is to consider the longer term effects (e.g. 12 months after the study) in both the exercise and control arms of RCTs in depression. Finally, the time period that we examined encompassed significant changes in the diagnostic criteria used to identify patients with MDD and this may have influenced the results.

Nevertheless, allowing for the aforementioned caveats, our results are consistent and indicate that research assessing the effectiveness of exercise in RCTs for those with depression face a challenge in competing with the powerful improvements within control groups. Similar debate has occurred regarding the rising placebo response to ‘inert pills’ in antidepressant RCTs (e.g. [17, 18, 80]). One previous meta-analysis of 29 RCTs [19] has investigated control group response in a non-pharmacological intervention in MDD (repetitive transcranial magnetic stimulation). The authors [19] found a similar increased control group improvement (SMD=-0.8) to ourselves but did not conduct robust meta-regression analyses. Thus, to our knowledge, our review provides the first comprehensive insight into the control group response in non-pharmacological interventions in depression.

A hypothesis for the larger control group response in inpatient trials might be that inpatients benefit from the multidisciplinary treatment (which often includes pharmacotherapy, psychotherapy and even electroconvulsive therapy) and the therapeutic milieu and intensive attention from health care providers, regardless of how frequent they exercise during their inpatient treatment. The greater control group response in people with a formal MDD diagnosis might be due to a higher baseline symptom severity in these patients. Also the percentage of drop-outs in the control group moderated the control group response. A recent correlational study [81] in the general population found that initial expectations predicted adherence to a 2-week walking program. Therefore, it is plausible that participants with lower expectations are more likely to drop-out from the study protocol than those who do expect a benefit. An alternative hypothesis might be that the supportive provision of empathy, a coherent narrative to understand one’s illness, and a therapeutic relationship are effective in themselves in the treatment of depression.

Ultimately, control groups responses are important inelucidating, explaining and optimizing the true antidepressant effects of regular physicalactivity in clinical settings. It is recommended that studies include real placebo groups in RCTs that are designed to examine the psychological effects of exercise. These could be traditional placebo (e.g. an inert placebo pill) or exercise placebo interventions, which is clearly a challenge. An exercise placebo intervention can be defined as an exercise intervention that is not generally recognized as efficacious, that lacks adequate evidence for efficacy, and that has no direct pharmacological, biochemical, or physical mechanism of action according to the current standard of knowledge [20]. An appropriate exercise placebo condition strategy might be to employ conditions that resemble some aspects of very low-intensity exercise, include equipment that passively moves the limbs of an individual (i.e. without an active muscle contraction) and imagery or hypnotic suggestion of exercise. It is however important that these kind of exercise interventions are administered in a context that is believable to participants. Future research trials should also assess the role of a-priori and subsequent expectations using validated questionnaires. Finally, a RCT is a limited research design for studying placebo effects in depression [82]. Due to the inability to blind participants to exercise training, expectations about the intervention are likely to introduce error to the observed effect of exercise [83] although this may increase the observed benefit of exercise also. A feasible alternative, that provides a better controlled estimate of the placebo effect, could be a between-subjects balanced-placebo design [20]. To date, no exercise studies in people with depression have attempted to use this design to study the size of placebo effects. Nevertheless, authors conducting exercise RCTs who find null effects, should carefully consider our results for the control group response, and future studies should consult our results in an attempt to minimize control group responses. Overall, to more sensitively test for an antidepressant effect of exercise, studies need to use control groups with a less pronounced response and/or adjust for factors which moderate this effect.

**Compliance with Ethical Standards**

**Conclusion**

We found consistent evidence that the control groups used in exercise studies experience large and significant improvements in depressive symptoms. Our results suggest that control groups responses play asubstantial role in the outcomes of studies investigating the impact of exercise forpeople with depression. The control group response in exercise studies is approximately double that reported in antidepressant meta-analyses.

**Funding**

Davy Vancampfort has support from the Research Foundation – Flanders (FWO-Vlaanderen). Brendon Stubbs was supported by the Collaboration for Leadership in Applied Health Research and Care South London theme for this article. No other sources of funding were used to assist in the preparation of this article.

**Conflicts of Interest**

Brendon Stubbs, Davy Vancampfort, Simon Rosenbaum, Philip Ward, Justin Richards, Michael Ussher and Felipe Schuch declare that they have no conflicts of interest relevant to the content of this review.

**Ethical Approval**

Ethical approval was not required to conduct this review.

**References**

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 386(9995):743-800.

2. Ferrari AJ, Charlson FJ, Norman RE, et al. Burden of depressive disorders by country, sex, age, and year: Findings from the Global Burden of Disease Study 2010. PLoS Med. 2013;10(11):1-12.

3. Warden D, Rush AJ, Trivedi MH, et al. The STAR\*D Project results: a comprehensive review of findings. Curr Psychiatry Rep. 2007;9(6):449-59.

4. Knapen J, Vancampfort D, Moriën Y, et al. Exercise therapy improves both mental and physical health in patients with major depression. Disabil Rehabil. 2015;37(16):1490-5.

5. Vancampfort D, Stubbs B, Ward PB, et al. Integrating physical activity as medicine in the care of people with severe mental illness. Aust N Z J Psychiatry. 2015;49(8):681-2.

6. Rethorst CD, Wipfli BM, Landers DM. The antidepressive effects of exercise: a meta-analysis of randomized trials. Sports Med. 2009;39(6):491-511.

7. Cleare A, Pariante CM, Young AH, et al. Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2008 British Association for Psychopharmacology guidelines. J Psychopharmacol. 2015;29(5):459-525.

8. Vancampfort D, Mitchell AJ, de Hert M, et al. Type 2 diabetes in patients with major depressive disorder: A meta‐analysis of prevalence estimates and predictors. Depress Anxiety. 2015;32(10):763-73.

9. Holt RIG, de Groot M, Lucki I, et al. NIDDK International conference report on diabetes and depression: Current understanding and future directions. Diabetes Care. 2014;37(8):2067-77.

10. Naci H, Ioannidis JPA. Comparative effectiveness of exercise and drug interventions on mortality outcomes: metaepidemiological study. BMJ. 2013;347:f5577.

11. Dugan SA, Bromberger JT, Eisuke S, et al. Association between physical activity and depressive symptoms: Midlife Women in SWAN. Med Sci Sports Exerc. 2015;47(2):335-42.

12. Schuch FB, Vasconcelos-Moreno MP, Borowsky C, et al. Exercise and severe major depression: Effect on symptom severity and quality of life at discharge in an inpatient cohort. J Psychiatr Res. 2015;61:25-32.

13. Chalder M, Wiles NJ, Campbell J, et al. Facilitated physical activity as a treatment for depressed adults: Randomised controlled trial. BMJ. 2012;344:e2758.

14. Krogh J, Videbech P, Thomsen C, et al. DEMO-II trial. Aerobic exercise versus stretching exercise in patients with major depression-a randomised clinical trial. PLoS One. 2012;7(10):e48316.

15. Papakostas GI, Østergaard SD, Iovieno N. The nature of placebo response in clinical studies of major depressive disorder. J Clin Psychiatry. 2015;76(4):456-66.

16. Schuch FB, Fleck MP. Is exercise an efficacious treatment for depression? A comment upon recent negative findings. FrontPsychiatry. 2013;4:20.

17. Rutherford BR, Mori S, Sneed JR, et al. Contribution of spontaneous improvement to placebo response in depression: A meta-analytic review. J Psychiatr Res  
. 2012;46(6):697-702.

18. Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. Am J Psychiatry. 2013;170(7):723-33.

19. Brunoni AR, Lopes M, Kaptchuk TJ, et al. Placebo response of non-pharmacological and pharmacological trials in major depression: a systematic review and meta-analysis. PLoS ONE. 2009;4(3):1-10.

20. Lindheimer J, O'Connor P, Dishman R. Quantifying the placebo effect in psychological 5 outcomes of exercise training: a meta-analysis of randomized trials. Sports Med. 2015;45(5):693-711.

21. Cooney GM, Dwan K, Greig CA, et al. Exercise for depression. Cochrane. 2013;9:CD004366.

22. Ekkekakis P. Honey, I shrunk the pooled SMD! Guide to critical appraisal of systematic reviews and meta-analyses using the Cochrane review on exercise for depression as example. Ment Health Phys Act. 2015;8:21-36.

23. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283(15):2008-12.

24. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. PLoS Clinical Trials. 2009;6(7):1-6.

25. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders – DSM-IV-TR. 4th edition. American Psychiatric Association; 2000.

26. World Health Organisation. The ICD-10 Classification of Mental and Behavioural Disorders – Diagnostic Criteria for Research. 1993.

27. Hamilton MAX. Development of a rating scale for primary depressive illness. Br J Soc Clin Psychol. 1967;6(4):278-96.

28. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. Arch Gen Psych. 1961:4; 561-71.

29. Silveira H, Moraes H, Oliveira N, et al. Physical exercise and clinically depressed patients: A systematic review and meta-analysis. Neuropsychobiology. 2013;67(2):61-8.

30. Josefsson T, Lindwall M, Archer T. Physical exercise intervention in depressive disorders: Meta-analysis and systematic review. Scand J Med Sci Sports. 2014;24(2):259-72.

31. Higgins JPT, S G. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).

32. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. BMJ. 2003;327(7414):557-60.

33. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics. 1994;50(4):1088-101.

34. Egger M, Davey SG, et al. Bias in meta-analysis detected by a simple, graphical test. BMJ. 1997;315(7109):629-34.

35. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics. 2000;56(2):455-63.

36. Rosenthal R. The file drawer problem and tolerance for null results. Psych Bulletin. 1979;86(3):638-41.

37. Blumenthal JA, Babyak MA, Doraiswamy PM, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. Psychosom Med. 2007 Sep-Oct;69(7):587-96.

38. Blumenthal JA, Babyak MA, O’Connor C, et al. Effects of exercise training on depressive symptoms in patients with chronic heart failure: The HF-ACTION randomized trial. JAMA. 2012;308(5):465-74.

39. Bonnet LH. Effects of Aerobic Exercise in Combination with Cognitive Therapy on Self Reported Depression [dissertation]. Hempstead, NY: Hofstra University,. 2005.

40. Brenes GA, Williamson JD, Messier SP, et al. Treatment of minor depression in older adults: a pilot study comparing sertraline and exercise. Aging Ment Health. 2007 Jan;11(1):61-8.

41. Chu I-H, Buckworth J, Kirby TE, et al. Effect of exercise intensity on depressive symptoms in women. Ment Health Phy Act. 2009;2(1):37-43.

42. Doyne EJ, Ossip-Klein DJ, Bowman ED, et al. Running versus weight lifting in the treatment of depression. J Consult Clin Psychol. 1987;55(5):748.

43. Dunn AL, Trivedi MH, Kampert JB, et al. Exercise treatment for depression: efficacy and dose response. Am J Prev Med. 2005 Jan;28(1):1-8.

44. Epstein D. Aerobic Activity Versus Group Cognitive Therapy: an Evaluative Study of Contrasting Interventions for the Alleviation of Clinical Depression. Reno: University of Nevada; 1986.

45. Foley LS, Prapavessis H, Osuch EA,et al. An examination of potential mechanisms for exercise as a treatment for depression: A pilot study. Ment Health Phys Act. 2008;1(2):69-73.

46. Fremont J, Craighead LW. Aerobic exercise and cognitive therapy in the treatment of dysphoric moods. Cogn Ther Res. 1987;11(2):241-51.

47. Gary RA, Dunbar SB, Higgins MK, et al. Combined exercise and cognitive behavioral therapy improves outcomes in patients with heart failure. J Psychosom Res. 2010;69(2):119-31.

48. Hemat-Far A, Shahsavari A, Roholla Mousavi S. Effects of selected aerobic exercises on the depression and concentrations of plasma serotonin in the depressed female students aged 18 to 25. J Appl Res Clin Exp Ther. 2012;12(1):47.

49. Hess-Homeier MJ. A comparison of Beck's cognitive therapy and jogging as treatments for depression; Missoula: University of Montana, 1981.

50. Hoffman JM, Bell KR, Powell JM, et al. A randomized controlled trial of exercise to improve mood after traumatic brain injury. PM&R. 2010 10//;2(10):911-9.

51. Klein MH, Greist JH, Gurman AS, et al. A comparative outcome study of group psychotherapy vs. exercise treatments for depression. Int J Ment Health. 1984:148-76.

52. Knubben K, Reischies FM, Adli M, et al. A randomised, controlled study on the effects of a short-term endurance training programme in patients with major depression. Br J Sports Med. 2007;41(1):29-33.

53. Krogh J, Saltin B, Gluud C, et al. The DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of strength versus aerobic versus relaxation training for patients with mild to moderate depression. J Clin Psychiatry. 2009 Jun;70(6):790-800.

54. Mather AS, Rodriguez C, Guthrie MF, et al. Effects of exercise on depressive symptoms in older adults with poorly responsive depressive disorder Randomised controlled trial. Br J Psychiatry. 2002;180(5):411-5.

55. McNeil JK, LeBlanc EM, Joyner M. The effect of exercise on depressive symptoms in the moderately depressed elderly. Psychol Aging. 1991;6(3):487.

56. Mota-Pereira J, Silverio J, Carvalho S, et al. Moderate exercise improves depression parameters in treatment-resistant patients with major depressive disorder. J Psychiatr Res. 2011 Aug;45(8):1005-11.

57. Mutrie N. Exercise as a treatment for depression within a national health service: Microform Publications, College of Human Development and Performance, University of Oregon; 1989.

58. Nabkasorn C, Miyai N, Sootmongkol A, et al. Effects of physical exercise on depression, neuroendocrine stress hormones and physiological fitness in adolescent females with depressive symptoms. Eur J Public Health. 2006;16(2):179-84.

59. Orth DK . Clinical Treatments for Depression. Morgantown: West Virginia University; 1979.

60. Setaro JL. Aerobic exercise and group counseling in the treatment of anxiety and depression: University Microfilms International/UMI; 1985.

61. Shahidi M, Mojtahed A, Modabbernia A, et al. Laughter yoga versus group exercise program in elderly depressed women: a randomized controlled trial. Int J Geriatr Psychiatry. 2011;26(3):322-7.

62. Sims J, Galea M, Taylor N, et al. Regenerate: assessing the feasibility of a strength‐training program to enhance the physical and mental health of chronic post stroke patients with depression. Int J Geriatr Psychiatry. 2009;24(1):76-83.

63. Singh NA, Stavrinos TM, Scarbek Y,. A randomized controlled trial of high versus low intensity weight training versus general practitioner care for clinical depression in older adults. J Gerontol A Biol Sci Med Sci. 2005 Jun;60(6):768-76.

64. Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of progressive resistance training in depressed elders. J Gerontol A Biol Sci Med Sci. 1997 Jan;52(1):M27-35.

65. Veale D, Le Fevre K, Pantelis C, et al. Aerobic exercise in the adjunctive treatment of depression: a randomized controlled trial. J R Soc Med. 1992 Sep;85(9):541-4.

66. Williams C, Tappen R. Exercise training for depressed older adults with Alzheimer's disease. Aging Ment Health. 2008;12(1):72-80.

67. Danielsson L, Papoulias I, Petersson EL, et al. Exercise or basic body awareness therapy as add-on treatment for major depression: A controlled study. J Affect Disord. 2014;168:98-106.

68. Kerling A, Tegtbur U, Gützlaff E, et al. Effects of adjunctive exercise on physiological and psychological parameters in depression: A randomized pilot trial. J Affect Disord. 2015;177:1-6.

69. Martiny K, Refsgaard E, Lund V, et al. A 9-week randomized trial comparing a chronotherapeutic intervention (wake and light therapy) to exercise in major depressive disorder patients treated with duloxetine. J Clin Psychiatry. 2012;73(9):1234.

70. Verrusio W, Andreozzi P, Marigliano B, et al. Exercise training and music therapy in elderly with depressive syndrome: A pilot study. Complement Ther Med. 2014;22(4):614-20.

71. Huang T-T, Liu C-B, Tsai Y-H, et al. Physical fitness exercise versus cognitive behavior therapy on reducing the depressive symptoms among community-dwelling elderly adults: A randomized controlled trial. Int J Nurs Stud. 2015;52(10):1542-52.

72. Salehi I, Hosseini SM, Haghighi M, et al. Electroconvulsive therapy and aerobic exercise training increased BDNF and ameliorated depressive symptoms in patients suffering from treatment-resistant major depressive disorder. J Psychiatr Res  
. 2014;57:117-24.

73. Hallgren M, Kraepelien M, Öjehagen A, et al. Physical exercise and internet-based cognitive behavioural therapy in the treatment of depression: randomised controlled trial. Br J Psychiatry. 2015:bjp. bp. 114.160101.

74. Belvederi Murri M, Amore M, Menchetti M, et al. Physical exercise for late-life major depression. Br J Psychiatry. 2015; 207(3):235-42.

75. Pilu A, Sorba M, Hardoy MC, et al. Efficacy of physical activity in the adjunctive treatment of major depressive disorders: preliminary results. Clin Pract Epidemiol Ment Health. 2007;3:8.

76. National Institute for Health and Care Excellence. Depression: Management of depression in primary and secondary care. National Institute for Health and Care Excellence, London.; 2004.

77. Nelson ME, Rejeski WJ, Blair SN, et al. Physical activity and public health in older adults: Recommendation from the American College of Sports Medicine and the American Heart Association. Medicine & Science in Sports & Exercise. 2007;39(8):1435-45.

78. Blair SN. Physical inactivity: the biggest public health problem of the 21st century. Br J Sports Med. 2009;43(1):1-2.

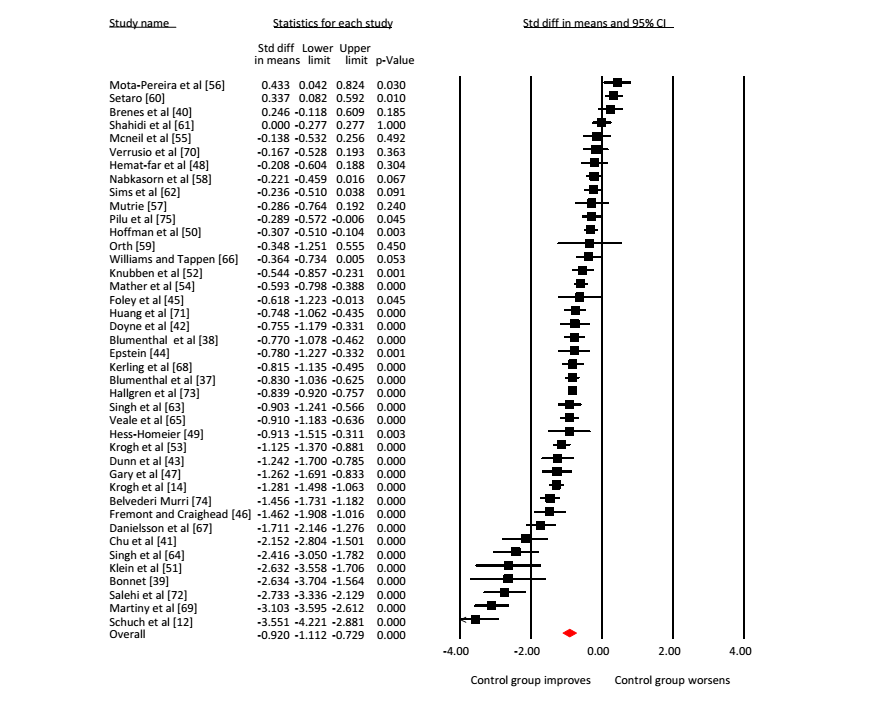
79. Bridge JA, Birmaher B, Iyengar S, et al. Placebo response in randomized controlled trials of antidepressants for pediatric major depressive disorder. Am J Psychiatry. 2009;166(1):42-9.

80. Stetler C. Adherence, expectations and the placebo response: Why is good adherence to an inert treatment beneficial? Psychol Health. 2014;29(2):127-40.

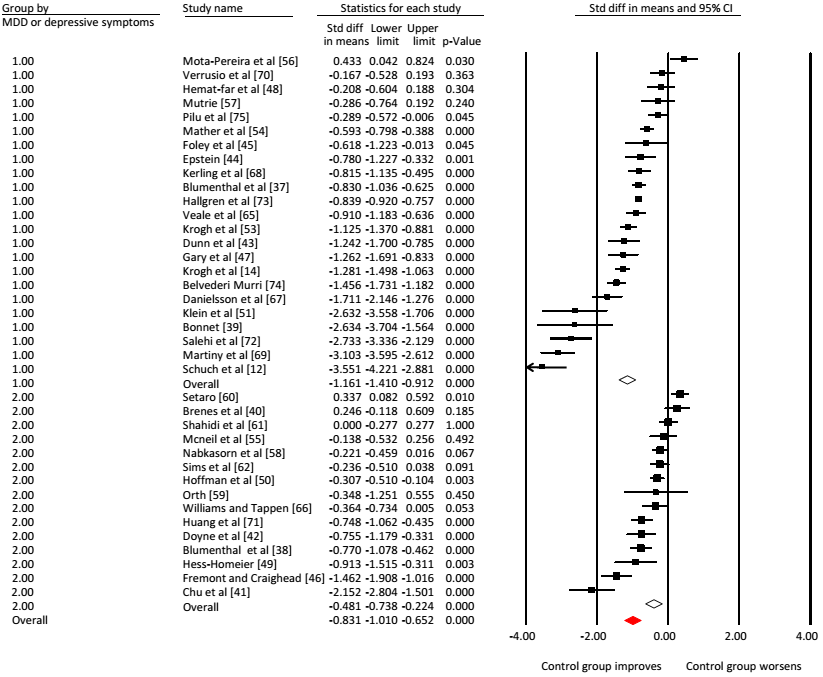
81. Benedetti F. Placebo effects: Understanding the mechanisms in health and disease. New York, NY, US: Oxford University Press; 2009.

82. Ojanen M. Can the true effects of exercise on psychological variables be separated from placebo effects? / Les effets veritables de l'exercice physique sur les variables psychologiques peuvent-ils etre separes des effets placebo ? Int J Sport Psychol. 1994;25(1):63-80.

**Figure 2 Meta-analysis of control group response in all studies.** CI = Confidence Interval, Std Diff = Standardized Difference.



**Figure 3 – Control group response in MDD (1.00=top) and depressive symptom (2.00=bottom) studies sub group analysis.** CI = Confidence Interval, MDD = Major Depressive Disorder, Std Diff = Standardized Difference.



**Table 1 Summary of included studies**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **CG Sample size** | **CG Age, y (mean)** | **CG % females** | **CG %antidepressant use** | **CG Depression diagnosis** | **CG Outcome** | **CG Strategy** | **Length of the trial (wk)** | **CG Drop out rate (%)** |
| Belvederi Murri et al [73] | 42 | 75.6 | 76 | 100 | MDD | HAM-D | Antidepressants | 24 | 28.5 |
| Blumenthal et al [36] | 49 | 52 | 77 | 0 | MDD | HAM-D | Placebo pill | 16 | 14.28 |
| Blumenthal et al [37] | 21 | 63.5 | 17 | 0 | Depressive symptoms | HAM-D | Placebo pill | 16 | 4.1 |
| Bonnet [38] | 6 | 25.33 | 33.3 | 0 | MDD | BDI-II | Psychotherapy | 6 | 33.3 |
| Brenes et al [39] | 12 | 73.9 | 50 | 0 | Depressive symptoms | HAM-D | TAU | 16 | ? |
| Chu et al [40] | 12 | 25.2 | 100 | 0 | Depressive symptoms | BDI-II | Stretching/low intensity exercise | 10 | 33 |
| Danielsson et al [66] | 20 | 46.3 | 80 | 100 | MDD | MADRS | TAU | 10 | 30 |
| Doyne et al [41] | 11 | 29.46 | 100 | 0 | Depressive symptoms | BDI | Wait list/NI | 8 | 13 |
| Dunn et al [42] | 13 | 34.5 | 62 | 0 | MDD | HAM-D | Stretching | 12 | 30.7 |
| Epstein [43] | 10 | ? | ? | ? | MDD | BDI | Wait list/NI | 8 | 0 |
| Foley et al [44] | 5 | ? | ? | 61.5 | MDD | BDI-II | Stretching/low intensity exercise | 12 | 61.53 |
| Fremont and Craighead [45] | 16 | ? | ? | 0 | Depressive symptoms | BDI | Psychotherapy | 10 | ? |
| Gary et al [46] | 15 | ? | ? | ? | MDD | HAM-D | TAU | 12 | 11.7 |
| Hallgren et al [72] | 312 | ? | ? | 24 | MDD | MADRS | TAU | 12 | 25.64 |
| Hemat-far et al [47] | 10 | ? | 100 | ? | MDD | BDI | Wait list/NI | 8 | 0 |
| Hess-Homeier [48] | 6 | 30 | ? | 0 | Depressive symptoms | BDI | Wait list/NI | 8 | 14.2 |
| Hoffman et al [49] | 39 | 37.1 | 50 | ? | Depressive symptoms | BDI | Wait list/NI | 10 | 2.5 |
| Huang et al [70] | 20 | 75.85 | 55 | 0 | Depressive symptoms | GDS-15 | Wait list/NI | 12 | 0 |
| Kerling et al [67] | 20 | 40.9 | 30 | 75 | MDD | MADRS | TAU | 6 | 0 |
| Klein et al [50] | 8 | 29.75 | 70.8 | 0 | MDD | SCL | Meditation | 12 | 17.4 |
| Knubben et al [51] | 18 | 50 | 55.5 | 88.8 | MDD + Bipolar | BRMS | Stretching/low intensity exercise | 10 daysa | 11.11 |
| Krogh et al [13] | 42 | 56.7 | 61.8 | 58.2 | MDD | HAM-D | Stretching/low intensity exercise | 16 | 30.9 |
| Krogh et al [52] | 59 | 43.4 | 67.2 | 15.1 | MDD | HAM-D | Stretching/low intensity exercise | 12 | ? |
| Martiny et al [68] | 37 | 46.9 | 64.9 | 100 | MDD | HAM-D | Wake up, bright light therapy | 9 | 10.8 |
| Mather et al [53] | 43 | 66.2 | 53 | 100 | MDD | HAM-D | TAU | 10 | 0 |
| Mcneil et al [54] | 10 | ? | ? | 0 | Depressive symptoms | BDI | Wait list/NI | 6 | 0 |
| Mota-Pereira et al [55] | 10 | 45.33 | 80 | 100 | MDD | HAM-D | TAU | 12 | 9.09 |
| Mutrie [56] | 7 | ? | ? | 0 | MDD | BDI | Wait list/NI | 4 | 36.3 |
| Nabkasorn et al [57] | 28 | 18.8 | 100 | 0 | Depressive symptoms | CES-D | Wait list/NI | 8 | 9.67 |
| Orth [58] | 2 | ? | ? | ? | Depressive symptoms | DACL | TAU | 4 | 0 |
| Pilu et al [74] | 20 | ? | 100 | 100 | MDD | HAM-D | TAU | 32 | 0 |
| Salehi et al [71] | 20 | 30.65 | 25 | 100 | MDD | HAM-D | ECT | 4 | 0 |
| Schuch et al [11] | 25 | 41.76 | 76 | Change during the trial | MDD | HAM-D | TAU | 3 | 4 |
| Setaro [59] | 25 | ? | ? | 0 | Depressive symptoms | MMPI | Wait list/NI | 10 | ? |
| Shahidi et al [60] | 20 | 68.4 | 100 | ? | Depressive symptoms | GDS | ? | ? | 16.6 |
| Sims et al [61] | 21 | 66.27 | 41 | ? | Depressive symptoms | PHQ-9 | TAU | 10 | 0 |
| Singh et al [62] | 15 | 72 | 53.3 | 0 | MDD + dysthymia | BDI | TAU | 10 | 0 |
| Singh et al [63] | 19 | 69 | 50 | 42 | MDD + dysthymia | HAM-D | TAU | 8 | 5 |
| Veale et al [64] | 29 | ? | ? | 34 | MDD | BDI | Wait list/NI | 12 | 17.2 |
| Verrusio et al [69] | 12 | 76.1 | 66.6 | 100 | MDD | GDS | TAU | 24 | 0 |
| Williams and Tappen [65] | 12 | ? | ? | ? | Depressive symptoms | CSDD | TAU | 16 | 9.09 |

**Key**: BDI = Beck Depression Inventory, BRMS = Bech-Rafaelsen Melancholia Scale, CES-D = Center for Epidemiologic Studies Depression Scale, CG = control group, CSDD = Cornel Scale for Depression in Dementia, DACL = Depression Adjective Check List, ECT = Electroconvulsotherapy, GDS = Geriatric Depression Scale, HAM-D = Hamilton Rating Scale for Depression, MADRS = Montgomery-Asberg Depression Rating Scale, MDD = Major Depressive Disorder, MMPI = Minnesota Multiphasic Personality Inventory, NI = No intervention, PHQ-9 = Patient Health Questionnaire, SCL = Symptom Checklist, TAU = Treatment as Usual. a = The intervention proposed by the Knubben [51] study had length of 10 days. "?" = Unclear.

**Table 2 Meta regression of moderators of control group response**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Moderator** | **Number RCTs** | **β** | **95% CI** | | **P value** | **R²** |
| **Main control group response** |  |  |  |  |  |  |  |
|  | Mean age (y) | 26 | 0.0226 | 0.0009 | 0.0461 | **0.05** | 0.13 |
|  | % females | 23 | -0.0089 | -0.0282 | 0.0104 | 0.36 | 0.04 |
|  | Year publication | 41 | -0.0094 | -0.036 | 0.0172 | 0.48 | 0.01 |
|  | % taking antidepressants | 30 | -0.0023 | -0.0099 | 0.0053 | 0.55 | 0.02 |
|  | Baseline depressive symptoms | 41 | -0.0058 | -0.0342 | 0.0226 | 0.69 | 0.00 |
|  | Duration control group | 40 | 0.0452 | -0.0049 | 0.0953 | 0.07 | 0.08 |
|  | Sample size | 41 | -0 | -0.0059 | 0.0058 | 0.98 | 0.0 |
|  | % drop out control group | 35 | -0.0233 | -0.0332 | -0.0134 | **0.001** | 0.16 |
|  | MDD diagnosis | 24 | -0.6319 | -1.2088 | -0.0549 | **0.03** | 0.13 |
|  | Depressive symptoms only | 14 | -0.7333 | -1.8073 | 0.3406 | 0.1 | 0.04 |
|  | **Multivariate meta regression (intercept)** | 24 | -1.3793 | -2.6614 | -0.0972 | **0.03** | **0.43** |
|  | MDD diagnosis |  | -0.5797 | -1.1943 | 0.00348 | 0.06 |  |
|  | Mean age (y) |  | 0.0175 | -0.0042 | 0.0392 | 0.11 |  |
|  | % drop out control |  | -0.0211 | -0.0388 | -0.0033 | **0.02** |  |
| **MDD Control group response** |  |  |  |  |  |  |  |
|  | Mean age | 14 | 0.0498 | 0.0108 | 0.0888 | **0.01** | 0.34 |
|  | % female | 15 | -0.0087 | -0.034 | 0.0165 | 0.49 | 0.04 |
|  | Year of publication | 23 | -0.0154 | -0.0545 | 0.0237 | 0.43 | 0.03 |
|  | % taking antidepressants | 15 | -0.0009 | -0.0115 | 0.0098 | 0.87 | 0.0 |
|  | Baseline depressive symptoms | 23 | -0.0005 | -0.0008 | 0.0001 | **0.008** | 0.01 |
|  | Duration control group | 23 | 0.0513 | 0.0147 | 0.0878 | **0.006** | 0.01 |
|  | Sample size | 23 | 0.0006 | -0.0040 | 0.0053 | 0.79 | 0.0 |
|  | % drop out control group | 21 | -0.0197 | -0.0367 | -0.0027 | **0.02** | 0.21 |
|  | **Multivariate meta regression** **(intercept)** | 14 | -1.9662 | -4.8796 | 0.9471 | 0.1 | **0.59** |
|  | Duration study |  | 0.0332 | -0.1089 | 0.1752 | 0.64 |  |
|  | Baseline depression |  | -0.029 | -0.0995 | 0.0414 | 0.41 |  |
|  | Mean age |  | 0.0221 | -0.0207 | 0.065 | 0.31 |  |
|  | % drop out control group |  | -0.0179 | -0.0366 | 0.0008 | **0.06** |  |
| **Depressive symptoms control group response** |  |  |  |  |  |  |  |
|  | Mean age | 9 | 0.0116 | -0.0033 | 0.0265 | 0.12 | 0.19 |
|  | Study year | 14 | 0.003 | -0.0245 | 0.0306 | 0.89 | 0.0 |
|  | Sample size | 14 | 0.0218 | -0.0136 | 0.0573 | 0.22 | 0.10 |
|  | Baseline depressive symptoms | 14 | 0.0111 | -0.0097 | 0.032 | 0.29 | 0.10 |

CI = Confidence Interval, MDD = Major Depressive Disorder, RCT = Randomized Clinical Trial

**Table 3: meta-analysis results of control group responses**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Analysis** | **Number of RCTs** | **Meta-analysis** | | | | **Heterogeneity** | | |
|  |  | **SMD** | **95% CI** | | **P value** | **I2** | **Q** | **P value** |
| **Main analysis** | 41 | -0.920 | -1.111 | -0.728 | **<0.0001** | 93 | 634 | <0.01 |
| Wait list/ TAU only | 26 | -0.642 | -0.853 | -0.431 | **<0.0001** | 92 | 332 | <0.01 |
| **Depression classification** |  |  |  |  |  |  |  |  |
| MDD | 24 | -1.161 | -1.410 | -0.912 | **<0.0001** | 93 | 354 | <0.01 |
| Depressive symptoms | 14 | -0.481 | -0.738 | -0.224 | **=0.001** | 88 | 120 | <0.01 |
| **Study quality** |  |  |  |  |  |  |  |  |
| High quality | 11 | -1.430 | -1.770 | -1.089 | **<0.0001** | 81 | 167 | <0.01 |
| Low quality | 30 | -0.713 | -0.925 | -0.498 | **<0.0001** | 90 | 318 | <0.01 |
| **Study setting** |  |  |  |  |  |  |  |  |
| Outpatient/ community | 36 | -0.884 | -1.048 | -0.639 | **<0.0001** | 91 | 462 | <0.01 |
| Inpatient | 4 | -1.783 | -2.40 | -1.161 | **<0.0001** | 85 | 91 | <0.01 |
| **Control group type** |  |  |  |  |  |  |  |  |
| Wait list/ no intervention | 11 | -0.429 | -0.679 | -0.179 | **<0.001** | 84 | 63 | <0.01 |
| Placebo pill | 2 | -0.800 | -1.464 | -0.137 | **=0.01** | 0 | 0.10 | 0.74 |
| Antidepressants | 1 | -1.456 | -2.399 | -0.513 | **=0.002** | 0 | 0 | 1 |
| Psychotherapies | 2 | -1.861 | -2.678 | -1.044 | **<0.0001** | 74 | 3 | 0.05 |
| Light intensity activity/ stretching | 6 | -1.134 | -1.541 | -0.727 | **<0.0001** | 81 | 27 | <0.01 |
| TAU/ usual care | 15 | -0.806 | -1.113 | -0.500 | **<0.0001** | 93 | 227 | <0.01 |
| Meditation | 1 | -2.631 | -3.924 | -1.339 | **<0.0001** | 0 | 0 | 1 |
| Light therapy/ Wake up therapy | 1 | -3.103 | -4.130 | -2.076 | **<0.0001** | 0 | 0 | 1 |
| ECT | 1 | -2.722 | -3.806 | -1.638 | **<0.0001** | 0 | 0 | 1 |
| **Comorbidities** |  |  |  |  |  |  |  |  |
| No major comorbidities | 35 | -0.978 | -1.188 | -0.769 | **<0.0001** | 85 | 590 | <0.01 |
| Included participants with comorbidities | 6 | -0.606 | -1.096 | -0.116 | **=0.01** | 74 | 21 | <0.01 |
| **Type of publication** |  |  |  |  |  |  |  |  |
| Peer review journal | 34 | -0.932 | -1.138 | -0.725 | **<0.0001** | 93 | 529 | <0.01 |
| Thesis | 7 | -0.849 | -1.341 | -0.357 | **<0.0001** | 91 | 81 | <0.01 |

**Key**: CI = Confidence Interval, MDD = Major Depressive Disorder, SMD= standardised mean difference, TAU = treatment as usual, RCT= randomised control trial, ECT = Electroconvulsotherapy