

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [editorial.bmjopen@bmj.com](mailto:editorial.bmjopen@bmj.com)

# BMJ Open

## Exercise for patients with major depression: a systematic review with meta-analysis and Trial Sequential Analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-014820
Article Type:	Research
Date Submitted by the Author:	19-Oct-2016
Complete List of Authors:	Krogh, Jesper; Kobenhavns Universitet Sundhedsvidenskabelige Fakultet, Mental Health Centre Copenhagen Speyer, Helene; University of Copenhagen Hjorthøj, Carsten; Copenhagen University Hospital, Mental Health Center Copenhagen Gluud, Christian; Copenhagen University Hospital Rigshospitalet, The Copenhagen Trial Unit, Centre for Clinical Intervention Research Nordentoft, Merete; University of Copenhagen, Mental Health Centre Copenhagen
<b>Primary Subject Heading</b>:	Mental health
Secondary Subject Heading:	Sports and exercise medicine, Evidence based practice
Keywords:	Exercise, MENTAL HEALTH, Systematic review, Meta-analysis, Evidence based medicine, Randomised clinical trials

SCHOLARONE™  
Manuscripts

Only

1  
2  
3  
4 **Exercise for patients with major depression: a systematic review with meta-**  
5 **analysis and Trial Sequential Analysis**  
6

7  
8 Jesper Krogh<sup>1</sup>, Carsten Hjorthøj<sup>1</sup>, Helene Speyer<sup>1</sup>, Christian Gluud<sup>1</sup>, Merete  
9 Nordentoft<sup>1</sup>  
10

11  
12  
13 <sup>1</sup> Mental Health Centre Copenhagen. Faculty of Health Sciences. University of Copenhagen, Copenhagen,  
14 Denmark.

15  
16 <sup>2</sup> Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University  
17 Hospital, Copenhagen, Denmark.  
18

19  
20  
21  
22  
23 **Corresponding author**

24 Jesper Krogh, [Jesper.krogh@dadlnet.dk](mailto:Jesper.krogh@dadlnet.dk)

25  
26 Mental Health Centre Copenhagen, Kildegårdsvej 28, Building 15. DK-2900, Hellerup, Denmark  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

### Objectives

To assess the benefits and harms of exercise in patients with depression.

### Design

Systematic review

### Data sources

Bibliographical databases were searched until the 17<sup>th</sup> of April, 2016.

### Eligibility criteria and outcomes

Eligible were randomised clinical trials assessing the effect of exercise in patients diagnosed with depression. Primary outcomes were depression severity, lack of remission, and serious adverse events. Secondary outcomes were quality of life and adverse events, as well as assessment of depression severity and lack of remission during follow-up after the intervention.

### Results

Thirty-one trials enrolling 2419 patients were included. The effect of exercise versus control on depression severity was -0.74 standardised mean difference (SMD) (95% CI -0.96 to -0.51;  $P < 0.001$ ; GRADE: very low quality). Restricting this analysis to the four trials that seemed less affected of bias, the effect vanished to -0.11 SMD (-0.41 to 0.18;  $P = 0.45$ ; GRADE: low quality). Exercise decreased the relative risk of no remission to 0.78 (0.68 to 0.90;  $P < 0.001$ ; GRADE: very low quality). Restricting this analysis to the two trials that seemed less affected of bias, the effect vanished to 0.95 (0.74 to 1.23;  $P = 0.78$ ). Trial Sequential Analysis excluded random error when all trials were analysed. Sub-group analyses found that trial size and intervention duration were inversely associated with effect size for both depression severity and lack of remission. There was no significant effect of exercise on secondary outcomes.

## Conclusions

Trials with less risk of bias suggested no antidepressant effects of exercise and there were no significant effects of exercise on quality of life, depression severity, or lack of remission during follow-up. Data for serious adverse events and adverse events was scarce not allowing conclusions for these outcomes.

## Systematic review registration

The protocol was published in the journal Systematic Reviews: 2015; 4:40

DOI: 10.1186/s13643-015-0030-6.

## Article Summary

### Strengths and limitations of this study

- The protocol for this review has previously been published
- Using meta-regression analysis, trial sequential analysis and the GRADE system the conclusions from this review is based on a firm and transparent platform
- Based on an extensive literature search, this review included 31 one trials allocating more than 2000 participants to exercise or control interventions
- All included participants were diagnosed with depression according to a diagnostic system
- Effect estimates from included trials had considerable heterogeneity

## Introduction

Depression is a common disease affecting up to 17% of the population during their lifetime.<sup>1,2</sup> Based on data from the World Health Organisation, depression is ranked as the second largest health-care problem globally, in terms of years lived with disability.<sup>3</sup> Depending on its severity, depression is often treated using psychotherapy, antidepressants, or a combination of both. However, the clinical benefits of antidepressants<sup>4,5</sup> and psychotherapy<sup>6-8</sup> has been challenged. Both treatments are costly in terms of time and money and may also have adverse effects. Compliance with antidepressant treatment is poor; the dropout rate in clinical trials is reported to be between 12% and 40% within the initial 6 to 8 weeks of treatment.<sup>4,9</sup>

The weakness of evidence for the beneficial effect of current interventions, along with problems related to low compliance and harms, has resulted in an interest in using alternative interventions. The use of exercise as an intervention has attracted considerable attention, and various forms of exercise varying in intensity have been assessed in a number of randomised clinical trials to test their effectiveness as a treatment for patients with depression. In 2011, we published a meta-analysis of randomised clinical trials examining the effect of exercise on depressive symptoms in patients with clinical depression.<sup>10</sup> The results suggested that referring patients with clinical depression to exercise programs was associated with a small to moderate effect on depressive symptoms. However, restricting the analysis to three trials with a low risk of bias, the effect estimate was non-significant. Since 2011, other reviews have been published on the effect of exercise on depressive symptoms,<sup>11</sup> in older people,<sup>12</sup> and in patients with chronic illnesses.<sup>13</sup> However, none of these reviews addressed the specific population of adults diagnosed with major depression according to valid diagnostic criteria, such as the International Classification of Diseases<sup>14</sup> or the Diagnostic and Statistical Manual of Mental Disorders.<sup>15</sup> The reviews contained a number of trials that included volunteers who were defined as being depressed on the basis of psychometric testing (for example, Beck

1  
2  
3  
4 Depression Inventory<sup>16</sup>), as opposed to individuals with a clinical diagnosis of major depression.

5  
6 Furthermore, several randomised clinical trials investigating the effect of exercise in clinically depressed  
7  
8 individuals have been published since our 2011 review.<sup>10</sup>  
9  
10

11  
12  
13  
14 The objectives of the present systematic review are to investigate the beneficial and harmful effects of  
15  
16 exercise, in terms of severity of depression, lack of remission, quality of life, and suicide versus controls  
17  
18 with or without co-interventions in adults with a clinical diagnosis of major depression. The current  
19  
20 systematic review differs from our previous review in a number of aspects.<sup>10</sup> We only considered trials  
21  
22 including participants diagnosed with depression according to a validated diagnostic system. We also  
23  
24 included trials including patients with somatic co-morbidity, e.g., cancer or diabetes. The harmful effects of  
25  
26 exercise interventions are also addressed, the intervention effects being assessed according to the grading  
27  
28 of recommendations assessment, development, and evaluation (GRADE) framework, and bibliographical  
29  
30 searches have been extended to include a Chinese and a South-American database until 2016.  
31  
32  
33  
34  
35  
36

### 37 **Methods/design**

38  
39 The protocol for this review has previously been published.<sup>17</sup>  
40  
41  
42  
43  
44

### 45 **Search strategy**

46  
47  
48 The following bibliographical databases was searched until the 17<sup>th</sup> of April, 2016: CENTRAL, MEDLINE,  
49  
50 EMBASE, Science Citation Index (Web of Science), LILACS, and Wanfang using medical subject headings  
51  
52 (MeSH or similar) when possible or text word terms: depression, depressive disorder and exercise, aerobic,  
53  
54 non-aerobic, physical activity, physical fitness, walking, jogging, running, bicycling, swimming, strength, or  
55  
56 resistance.  
57  
58  
59  
60



## Trial selection

One investigator (JK) examined titles and abstracts to remove obviously irrelevant reports. Two investigators (JK + HS) examined full text reports and abstracts determining compliance with inclusion criteria. A trial was considered eligible if it was a randomised clinical trials including patients diagnosed as having major depression according to a valid and recognised diagnostic system (that is, Research Diagnostic Criteria (RDC),<sup>18</sup> International Classification of Diseases (ICD),<sup>14</sup> or Diagnostic and Statistical Manual of Mental disorders (DSM)<sup>15</sup>) and included participants aged >17 years. Both abstracts and full text reports included.

Trials were excluded if they measured depression immediately after a single bout of exercise, compared one form of exercise versus another, or compared different exercise intensities without including a control group. The trials had to allocate participants to an exercise intervention versus a control group (that is, exercise versus a control group receiving no intervention or treatment as usual or an attention control using light exercise) or using exercise as an add-on-treatment (that is, exercise plus usual treatment in the experimental group versus usual treatment alone in the control group). Exercise intervention was defined as a systematic physical intervention with the intention to increase muscle strength and/or cardiovascular fitness, e.g., running, swimming or weight lifting. In case of attention control, it should specifically be mentioned by the authors that the intervention was intended as a control intervention.

## Outcomes

The primary outcomes were 1) depressive symptoms measured on a continuous scale assessed at the end of the intervention; 2) lack of remission, that is, a binary outcome of the proportion of participants in each

1  
2  
3  
4 intervention group of the trial who did not obtain remission at the end of the intervention according to the  
5 authors' own definition; and 3) serious adverse events defined according to ICH-GCP as any untoward  
6 medical occurrence that was life threatening, resulted in death or persistent or significant disability (ICH-  
7 GCP 1997).<sup>19</sup> Serious adverse events accordingly include suicide attempts as well as suicides. The secondary  
8 outcomes were quality of life, non-serious adverse events, as well as depressive symptoms and lack of  
9 remission assessed after the intervention.  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19

### 20 **Data extraction**

21  
22  
23 Two authors (JK, HS) independently extracted data using a pre-piloted structured form. Any discrepancies  
24 in the data extraction or inclusion/exclusion of trials was resolved by referring to the original papers. CG or  
25 MN assisted as adjudicator in cases of disagreements. Data extraction included, in addition to outcomes,  
26 information regarding country of origin, number of randomised participants, number of participants  
27 included in efficacy analysis, mean age of participants, diagnostic system, baseline assessment of  
28 depression severity, type of intervention, frequency of intervention, and duration of intervention. JK and  
29 CH independently performed the assessment of bias domains. The authors JK, CG, and MN have previously  
30 published trial reports assessing the effect of exercise in patients with depression,<sup>20;21</sup> and to reduce the  
31 risk of academic bias two additional authors were included in the current systematic review (CH, HS).  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

### 46 **Risk of bias assessment**

47  
48  
49 Definitions in the assessment of bias risk of a trial was conducted according to the Cochrane Handbook for  
50 Systematic Reviews of Interventions<sup>22</sup> of the following domains: allocation sequence generation, allocation  
51 concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome  
52 data, selective outcome reporting, for-profit bias, and other bias. Trials assessed as having 'low risk of bias'  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 in all of the above specified domains were considered 'trials at low risk of bias'. Trials assessed as having  
5  
6 'uncertain risk of bias' or 'high risk of bias' in one or more of the above specified domains were considered  
7  
8 trials with 'high risk of bias'. In line with our previous systematic review<sup>10</sup> and the latest Cochrane review on  
9  
10 exercise for depression,<sup>23</sup> trials at low risk of bias in the allocation concealment domain, blinded outcome  
11  
12 assessment domain, and the incomplete outcome data domain were characterised as 'trials potentially  
13  
14 having less risk of bias than other trials at high risk of bias'. Trials assessing the effect of behavioural  
15  
16 interventions are rarely able to mask the allocation, and participants and health care providers are  
17  
18 therefore not blinded. Therefore, we will also report the number of trials at low risk of bias in the remaining  
19  
20 domains.  
21  
22  
23  
24  
25  
26

### 27 **Data synthesis and analysis**

28  
29 In order to be able to include all of the trials in our meta-analysis, estimates of standardised mean  
30  
31 difference (SMD) for each individual trial was carried out. SMD is the mean difference in depression score  
32  
33 between the exercise and control groups divided by the pooled standard deviation. The result is a unit free  
34  
35 effect size. By convention, SMD effect sizes of 0.2, 0.5 and 0.8 are considered small, medium and large  
36  
37 intervention effects. In case post-test scores as well as change from baseline was reported, post-test scores  
38  
39 were preferred. For dichotomous variables, we calculated the risk ratio (RR) with a 95% confidence interval.  
40  
41 It was expected that some trials would have several intervention groups. Data from the experimental  
42  
43 groups was pooled and compared with the data from the control group. In case of discrepancies between  
44  
45 the random-effects model analysis and the fixed-effect model analysis, both results are reported;  
46  
47 otherwise, only results from the random-effects analysis is reported. The degree of heterogeneity was  
48  
49 quantified using the I-squared statistic,<sup>24</sup> which can be interpreted as the percentage of variation observed  
50  
51 between the trials attributable to between-trial differences, rather than sampling error (chance).  
52  
53  
54  
55 Heterogeneity was explored by analyses of sub-groups (see below).  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7 For the primary outcomes, Trial Sequential Analysis was performed.<sup>25;26</sup> In order to calculate the required  
8 information size and the cumulative Z-curve's eventual breach of relevant trial sequential monitoring  
9 boundaries, the required information size for the primary continuous outcome was based on type I error of  
10 5%, a beta of 10%, the standard error of the meta-analysis, and a minimal difference of three points on the  
11 HAM-D<sub>17</sub>.<sup>17</sup> In order to calculate the required information size and the cumulative Z-curve's eventual breach  
12 of relevant trial sequential monitoring boundaries, the required information size for lack of remission was  
13 based on type I error of 5%, a beta of 10%, the proportion of patients in the control group with the  
14 outcome, and a relative risk reduction of 15% and 30%.  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

27 Bayes factors were calculated for all primary outcomes.<sup>27</sup> Low P-values suggest that we can reject the null-  
28 hypothesis. But even a low P-value from a meta-analysis can be misleading if there is also a low probability  
29 that data are compatible with the anticipated intervention effect. In other words, the probability that the  
30 actual measured difference in effect of the compared interventions resulted from an a priori anticipated  
31 'true' difference needs to be considered. For this purpose, it is helpful to calculate the Bayes factor, which  
32 is the ratio of the P-value probabilities of the meta-analysis result divided by the probability of the  
33 anticipated effect, or 'true' effect.<sup>27</sup> As suggested by Jakobsen et al.,<sup>27</sup> a Bayes factor lower than 0.1  
34 together with a low P-value suggest, if bias can be ruled out, that the observed result is compatible with the  
35 a priori expected effect. If the Bayes factor is higher than 0.1 the result is not compatible with the a priori  
36 expected effect and the effect may be lower.  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

53 To assess the potential impact of missing data (incomplete outcome data bias) we did sensitivity analysis of  
54 missing data using the following strategy: a 'best-worst' case scenario was assessed, assuming that all  
55 participants lost to follow-up in the intervention group had a beneficial outcome (the group mean minus 1  
56  
57  
58  
59  
60

1  
2  
3  
4 standard deviation (SD)), and all those with missing outcomes in the placebo group have had a harmful  
5  
6 outcome (the group mean plus 1 SD and 2 SD). In addition, the reverse 'worst-best-case' scenario analysis  
7  
8 was also performed.<sup>27</sup> Missing data for the 'lack of remission' outcome was imputed in sensitivity analysis  
9  
10 according to the following scenarios:<sup>28</sup> 1) poor outcome analysis: assuming that all of the drop-  
11  
12 outs/participants lost from both the experimental and the control arms experienced the outcome, including  
13  
14 all randomised participants in the denominator; 2) good outcome analysis: assuming that none of the drop-  
15  
16 outs/participants lost from the experimental and the control arms experienced the outcome, including all  
17  
18 randomised participants in the denominator; 3) extreme case analysis favouring the experimental  
19  
20 intervention ('best-worse' case scenario): none of the drop-outs/participants lost from the experimental  
21  
22 arm, but all of the drop-outs/participants lost from the control arm experienced the outcome, including all  
23  
24 randomised participants in the denominator; and 4) extreme case analysis favouring the control ('worst-  
25  
26 best' case scenario): all of the drop-outs/participants lost from the experimental arm, but none from the  
27  
28 control arm experienced the outcome, including all randomised participants in the denominator.  
29  
30  
31  
32  
33  
34  
35

### 36 Subgroup analyses

37  
38 In subgroup analyses, the possible effects of variables on intervention effects on outcomes and  
39  
40 heterogeneity were compared. Trials potentially having less risk of bias (i.e., trials with adequate allocation  
41  
42 concealment, blinded outcome assessment, and intention to treat analysis) were compared to trials at high  
43  
44 risk of bias. The effect of age was assessed by comparing trials including older participants (mean age >59  
45  
46 years) to trials including younger participants (mean age <60 years). The effect of type of exercise was  
47  
48 assessed by comparing trials using group exercises compared to trials using individual exercise. The effect  
49  
50 of duration of intervention was assessed by comparing trials with short duration of intervention to trials  
51  
52 with long duration of intervention splitting by the median time of duration. The effect of type of control  
53  
54 group was assessed by comparing trials using attention control to trials with waitlist controls and  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 comparing trials with exercise as add-on to medication to trials not using any medication. In addition, a  
5  
6 within-study comparison of low-dose exercise versus high-dose exercise in trials using different exercise  
7  
8 intensities was performed. The effect of co-morbid somatic disease was assessed by comparing the effect  
9  
10 estimates from trials including patients with depression compared to trials including patients with  
11  
12 depression in addition to a somatic disease. Publication bias was assessed by visual inspection of a funnel  
13  
14 plot and by Egger's test and if publication bias plausible Duval's and Tweedie's trim and fill procedure was  
15  
16 conducted.<sup>29</sup>  
17  
18  
19  
20

21  
22 We assessed and graded the evidence according to the grading of recommendations assessment,  
23  
24 development, and evaluation (GRADE) for high risk of bias, imprecision, indirectness, heterogeneity, and  
25  
26 publication bias.<sup>30</sup> Based on this assessment, the intervention is graded accordingly: 'high quality'- we are  
27  
28 very confident that the true effect lies close to that of the estimate of the effect; 'moderate quality'- we are  
29  
30 moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the  
31  
32 effect, but there is a possibility that it is substantially different; 'low quality'- our confidence in the effect  
33  
34 estimate is limited: the true effect may be substantially different from the estimate of the effect; 'very low  
35  
36 quality'- we have very little confidence in the effect estimate: the true effect is likely to be substantially  
37  
38 different from the estimate of the effect.<sup>31</sup>  
39  
40  
41  
42  
43  
44  
45

#### 46 **Deviations from our protocol**

47  
48 Post-hoc we included trials using the Chinese Classification of Mental Disorders (CCMD) as well as a few  
49  
50 trials including patients classified as having 'minor depression'. The CCMD system closely adhere to the ICD  
51  
52 and DSM systems and have been found highly compatible in field studies, so these studies were included.<sup>32</sup>  
53  
54

55 A few trials included some patients classified as having 'minor depression' according to the trials chosen  
56  
57 diagnostic system (e.g., DSM), and it is questionable if these patients have major depression. We therefore  
58  
59  
60

1  
2  
3  
4 decided to include these trials but also to conduct a sub-group analysis exclusively including patients with  
5  
6 major depression. Post-hoc we also included a sub-group analysis according to trial size. Trials were divided  
7  
8 into small or large trials using the median of total n included in the efficacy analysis. We did not conduct  
9  
10 Trial Sequential Analysis based on a relative risk reduction of 30% of lack of remission as this was an  
11  
12 implausible effect.  
13

### 14 15 16 17 18 **Patient involvement**

19  
20  
21 Depressed patients were not involved in this study.  
22  
23  
24  
25  
26

### 27 **Results**

#### 28 29 **Bibliographical search and trial characteristics**

30  
31  
32 The main bibliographical search was conducted the 26<sup>th</sup> of August, 2015 and the final updates were  
33  
34 conducted on the 17<sup>th</sup> of April, 2016. As illustrated in Figure S1, we identified 40 publications reporting the  
35  
36 effect of exercise on depressive symptoms in 31 randomised clinical trials.<sup>20;21;33-71</sup> Four-teen trials were  
37  
38 conducted in Europe,<sup>20;21;38;47;50;51;53;59;63-66;72;73</sup> seven in the U.S.A.,<sup>36;37;41;43;58;62;74</sup> six in Asia,<sup>45;67-71</sup> two in  
39  
40 Australia,<sup>52;56</sup> and two in South-America.<sup>54;61</sup> A total of 2,419 patients were randomised and 2,331 were  
41  
42 included in the efficacy analysis of benefit. 10 trials included inpatients<sup>45;47;54;65;67-72</sup> and five trials included  
43  
44 participants with a mean age above 60 years.<sup>50;52;56;58;59</sup> No trials exclusively included patients with  
45  
46 comorbid somatic disease. Please see Table 1 for trial characteristics.  
47  
48  
49  
50  
51  
52

#### 53 *Bias risk assessment*

54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 Sequence generation was adequate in 12/31 (39%), allocation concealment was adequate in 12/31 (39%)  
5  
6 trials, blinding of participants and trial personnel was adequate in 0/31 (0%), blinded outcome assessment  
7  
8 was performed in 16/31 (52%), low risk of bias in the 'incomplete outcome data' domain was found in  
9  
10 12/31 (39%) trials, selective outcome reporting domain was adequate in 27/31 (87%), for profit bias  
11  
12 domain was adequate in 15/31 (48%) and 21/31 (68%) were free of other bias. All trials were at high risk of  
13  
14 bias. Given the nature of the intervention, no trial had blinded participants or trial personnel, however, two  
15  
16 trials had low risk of bias in all other bias domains.<sup>21;52</sup> Five trials (16%) were sponsored by for profit  
17  
18 organisations: three trials were supported by pharmaceutical companies,<sup>51;72;75</sup> one trial by a company  
19  
20 producing fitness machines,<sup>43</sup> and one trial by an insurance company.<sup>20</sup> According to our a priori defined  
21  
22 criteria, 4/31 (13%) trials potentially had less risk of bias than the other trials at high risk of bias.<sup>20;21;52;54</sup>  
23  
24  
25 Please see Table 2 for details on assessment of risks of bias.  
26  
27  
28  
29  
30

### 31 **Primary outcomes**

#### 32 **The effect of exercise on depression severity**

33  
34  
35 All included trials provided a continuous outcome on depression severity for the assessment of the exercise  
36  
37 intervention encompassing 2,331/2,419 randomised patients (96.4%). The effect of intervention versus  
38  
39 control was a standardised mean difference (SMD) of -0.74 (95% CI -0.96 to -0.51; P<0.001) (Figure 1.). This  
40  
41 corresponds to an effect on the HAM-D<sub>17</sub> scale of -4.6 (95% CI -6.0 to -3.2) points.  
42  
43  
44  
45  
46  
47

#### 48 *Missing data*

49  
50  
51 Missing outcome analysis for depression as a continuous outcome did not markedly change the effect  
52  
53 estimates. The least favourable outcome for the exercise intervention was the worse/best outcome  
54  
55 analysis using +2 SD resulting in an effect estimate of -0.61 SMD (95% CI -0.84 to -0.37; P<0.001) (Table S1).  
56  
57  
58  
59  
60



### *Heterogeneity and subgroup analysis*

The  $I^2$  was 82% suggesting substantial heterogeneity. Subgroup analysis revealed that the effect estimates for trials potentially having less risk of bias was -0.11 SMD (95% CI. -0.41 to 0.18;  $P = 0.45$ ;  $I^2 = 62\%$ ) compared to that of the trials at high risk of bias -0.85 SMD (95% CI. -1.10 to -0.60;  $P < 0.001$ ;  $I^2 = 82\%$ ) (test of sub-group difference,  $P = 0.0002$ ). In addition, trials including 52 patients or less had a pooled estimate of -1.30 SMD (95% CI -1.74 to -0.86;  $P < 0.001$ ;  $I^2 = 77\%$ ) compared to that of larger trials of -0.40 SMD (95% CI -0.60 to -0.19;  $P < 0.001$ ;  $I^2 = 76\%$ ) (test of sub-group difference,  $P < 0.001$ ). Trials of short duration of intervention (less than 10 weeks) had a SMD of -0.93 (95% CI -1.11 to -0.88;  $P < 0.001$ ;  $I^2 = 19\%$ ) compared to trials with longer duration of intervention, -0.58 SMD (95% CI -0.88 to -0.28;  $P < 0.001$ ;  $I^2 = 86\%$ ) (test of sub-group difference,  $P = 0.05$ ). Effect estimates from trials including patients with minor depression compared to trials exclusively including patient with major depression did not differ (test of sub-group difference,  $P = 0.67$ ).

Four trials allocated 206 patients to different exercise intensities/doses.<sup>43;56;71;76</sup> Comparing the post-intervention depression scores for patients allocated to either high intensity/high dose or low intensity/low dose exercise showed a difference of -0.40 SMD (95% CI -0.67 to -0.12;  $P=0.005$ ;  $I^2 = 0\%$ ) in favour of high intensity/high dose exercise. As shown in Table 3, no other trial characteristic significantly explained any of the observed heterogeneity. Please see Table S2 for trial characteristics used to explore heterogeneity.

### *Trial Sequential Analysis and diversity adjusted required information size*

The diversity adjusted required information size for HAM-D<sub>17</sub> as a continuous outcome was calculated based on our anticipated intervention effect of a minimal relevant difference of 3.0 HDRS points, a standard

1  
2  
3  
4 deviation of 6.78 points, a risk of type I error of 0.05, a power of 90% and the observed diversity of 92% to  
5  
6 2610 participants. Only 14 trials reported results from HAM-D<sub>17</sub><sup>20;21;36;37;41;42;50;51;53;54;56;66;68;76</sup> with an accrued  
7  
8 1124 participants. As shown in Figure S2, the cumulative Z-curve just crossed the trial sequential  
9  
10 monitoring boundary for benefit. With the aforementioned settings, the pooled estimate is therefore less  
11  
12 likely to be a random finding due to lack of power or multiple testing if bias could be ignored.  
13  
14  
15  
16  
17  
18  
19  
20

### 21 *Bayes factor*

22  
23  
24 Fourteen trials reported effect estimates using the HAM-D<sub>17</sub><sup>20;21;36;37;41;43;50;51;53;61;66;68;76;77</sup> Based on these  
25  
26 trials, Bayes factor was calculated ( $\delta = -3.37$ ;  $SE_{\delta} = 0.96$ ;  $\mu_a = -3.0$ ) and was found to be 0.002, which is  
27  
28 below the Bayes factor threshold for significance of 0.1, supporting the intervention effect if bias could be  
29  
30 ignored.  
31  
32  
33  
34  
35

### 36 *Publication bias*

37  
38 Inspection of the funnel-plot (not shown) suggested that small trials with small or no effect of exercise  
39  
40 were missing. Egger's test supported the suspicion of publication bias,  $P < 0.00001$ . Using the Duval and  
41  
42 Tweedie's trim and fill procedure, the estimate was reduced to -0.28 SMD (95% CI -0.52 to -0.04). This  
43  
44 corresponds to an effect on the HAM-D<sub>17</sub> scale of -1.8 (95% CI -3.2 to -0.25).  
45  
46  
47  
48  
49  
50

### 51 **The effect of exercise on depression – lack of remission**

52  
53 Nineteen trials, randomising 1825 patients and including 1639 patients (90%) in final analysis reported  
54  
55 remission as an outcome.<sup>20;21;36-38;41;43;45;47;51;52;54;58;59;63;66-68;70</sup> The RR for lack of remission was 0.78 (95% CI  
56  
57  
58  
59  
60

1  
2  
3  
4 0.68 to 0.90; P=0.0008) in favour of the intervention using a random-effects analysis. The  $I^2$  was 69%  
5  
6 suggesting substantial heterogeneity. The forest plot for the intervention effect on lack of remission is  
7  
8 illustrated in Figure S3.  
9

#### 10 11 12 13 14 *Missing data*

15  
16  
17 The scenario in least favour of the intervention was the 'poor' outcome analysis having an effect estimate  
18 of RR 0.88 (95% CI 0.83 to 0.94) P=0.0002;  $I^2$  = 69%. As shown in Table S1, the remaining scenarios did not  
19  
20 substantially differ from the main analysis.  
21  
22

#### 23 24 25 26 *Heterogeneity and subgroup analysis*

27  
28  $I^2$  was 69% for the outcome lack of remission suggesting substantial heterogeneity. For this outcome, only  
29 two trials<sup>21;77</sup> were considered as trials potentially having less risk of bias than the other trials at high risk of  
30 bias. The RR of these two trials was 0.95 (95% CI 0.74 to 1.23; P=0.78) compared to 0.77 (96% CI 0.64 to  
31 0.92; P=0.003) for trials at high risk of bias, test of subgroup difference, P=0.19). Trials including 52  
32 participants or less in their final analysis had a RR 0.62 (95% CI 0.50 to 0.76; P<0.001;  $I^2$  = 45%) compared to  
33 0.95 (95% CI 0.80 to 1.12; P=0.52;  $I^2$  = 68%) for larger trials (test of sub-group difference, P=0.002). Also,  
34 trials with a duration of less than 10 weeks had a RR of 0.63 (95% CI 0.51 to 0.77; P<0.001;  $I^2$  = 40%)  
35 compared to 0.93 (95% CI 0.78 to 1.10; P=0.39;  $I^2$  = 69%) for trials of a longer duration (test of sub-group  
36 difference, P=0.004). As shown in Table S3, no other trial characteristic significantly explained any of the  
37 observed heterogeneity. Please see Table S2 for trial characteristics used to explore heterogeneity.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

#### 53 *Trial Sequential Analysis and diversity adjusted required information size*

54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 The diversity adjusted required information size for lack of remission was calculated based on our observed  
5  
6 diversity of 74%, a proportion in the control group with lack of remission of 66%, an anticipated  
7  
8 intervention effect of 15% relative risk reduction, a risk of type I error of 0.05% and a power of 90%. As  
9  
10 shown in Figure S4, the cumulative Z curve just crossed the trial sequential monitoring boundary for  
11  
12 benefit. With the aforementioned settings, the pooled estimate is therefore less likely to be a random  
13  
14 finding due to lack of power or multiple testing if bias could be ignored.  
15  
16  
17  
18  
19  
20  
21  
22  
23

#### 24 *Bayes factor*

25  
26 Bayes factor was calculate based on the observed relative risk of remission, the associated standard error,  
27  
28 and an anticipated intervention effect of relative increase in number of patients with remission by 15% ( $\delta =$   
29  
30  $-0.248$ ;  $SE_{\delta} = 0.08$ ;  $\mu_{\delta} = -0.163$ ). Bayes factor was 0.02, which is below the Bayes factor threshold for  
31  
32 significance of 0.1.  
33  
34  
35  
36  
37  
38

#### 39 *Publication bias*

40  
41 Inspection of the funnel-plot (not shown) suggested that small trials with small or no effect of exercise  
42  
43 were missing. Egger's test supported the suspicion of publication bias,  $P=0.002$ . Imputing theoretically  
44  
45 missing studies by the Duval and Tweedie's trim and fill procedure, reduced the estimate of intervention  
46  
47 effect to a relative risk reduction of 0.93 (95% CI 0.79 to 1.11).  
48  
49  
50  
51  
52  
53

#### 54 **The effect of exercise on serious adverse events**

1  
2  
3  
4 Serious adverse events (i.e., death or suicide attempts) were reported in only three trials.<sup>20;21;56</sup> In these  
5  
6 trials, one suicide attempt<sup>21</sup> and one death by suicide<sup>20</sup> were recorded in the intervention groups. The RR  
7  
8 for death or suicide in the two trials was 2.21 (95% CI 0.24 to 20.21; P=0.48; I<sup>2</sup> = 0%) as illustrated in Figure  
9  
10 S5.  
11

### 12 13 14 15 16 *Missing data*

17  
18 Missing outcome analysis for 'serious adverse events' varied according to missing data scenario: poor  
19  
20 outcome analysis relative risk, 0.92 (95% CI 0.37 to 2.30; P=0.86; I<sup>2</sup> = 60.0%), good outcome analysis, 2.19  
21  
22 (95% CI 0.23 to 20.76; P=0.50; I<sup>2</sup> = 0.0%), best/worst outcome analysis – 0.08 (95% CI 0.02 to 0.34; P=0.001;  
23  
24 I<sup>2</sup> = 5.4%), worst/best outcome analysis 19.17 (95% CI 2.64 to 139.2; P=0.004; I<sup>2</sup> = 0.0%).  
25  
26  
27  
28  
29  
30

### 31 *Trial Sequential Analysis and Bayes analysis*

32  
33 We decided not to conduct Trial Sequential Analysis or Bayes analysis due to too sparse data.  
34  
35  
36  
37  
38

### 39 *Publication bias*

40  
41  
42 Only 3/31 trials reported on this outcome and no formal assessment for publication bias was made.  
43  
44 However, the lack of reporting in the vast majority of trials suggest risk publication bias.  
45  
46  
47  
48

### 49 **Secondary outcomes**

50  
51  
52 *The effect of exercise on quality of life*  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 Eight trials randomising 901 participants reported on quality of life,<sup>20;21;36;38;54;58;69;78</sup> observing that patients  
5  
6 allocated to exercise did not have significantly better quality of life (SMD 0.43; 95% CI -0.04 to 0.91;  
7  
8 P=0.08). The  $I^2$  was 89% showing substantial heterogeneity (Figure S6).  
9  
10

#### 11 12 13 14 *Non-serious adverse events*

15  
16 Non-serious adverse events were reported in only nine trials.<sup>20;21;37;54;56;58;63;65;66</sup> Five trials reported on  
17  
18 musculoskeletal adverse events without conducting formal tests<sup>56;58;63;65;66</sup> and four trials reported on  
19  
20 number of patients with high depression scores post-intervention compared to baseline  
21  
22 assessment.<sup>20;21;63;66</sup> The RR for increased severity of depression post-intervention was 0.83 (95% CI 0.40 to  
23  
24 1.70; P=0.60;  $I^2 = 0.0\%$ ).  
25  
26  
27  
28  
29  
30

#### 31 32 33 34 *The effect of exercise on depression beyond the duration of the intervention*

35  
36 Assessment of depression beyond the intervention was conducted in seven trials,<sup>20;36;38;50;58;61;79</sup> with a  
37  
38 median duration between end of intervention and assessment of depression of 6 months (range 5 to 23.5  
39  
40 months). The SMD between the intervention group and the control group using a random effects analysis  
41  
42 was -0.10 (95% CI -0.28 to 0.09; P=0.31;  $I^2 = 19.5\%$ ). The  $I^2$  for this estimate was 19.5% suggesting low  
43  
44 heterogeneity (See Figure S7).  
45  
46  
47

48  
49 Remission beyond the intervention was assessed in five trials,<sup>20;36-38;52</sup> and the relative risk of lack of  
50  
51 remission was 0.95 (95% CI 0.82 to 1.11; P=0.53) with an  $I^2$  of 0.0% (See Figure S8).  
52  
53  
54  
55

#### 56 57 58 59 60 *GRADE assessments*

1  
2  
3  
4 The GRADE assessments are presented in Table 4, and quality of evidence for both primary and secondary  
5  
6 outcomes was very low or low.  
7  
8

## 9 10 11 **Discussion**

12  
13  
14 Thirty-one clinical trials allocating more than 2400 participants diagnosed with depression according to  
15  
16 validated diagnostic instruments were included in the present systematic review. Pooled estimates  
17  
18 suggested moderate antidepressant effect assessed both as a continuous outcome and as lack of remission.  
19  
20 Due to risk of bias, inconsistency of effect estimates, and publication bias we have, however, very little  
21  
22 confidence in these effect estimates. Subgroup analyses exploring reasons for the heterogeneity found that  
23  
24 trials potentially having less risk of bias than other trials at high risk of bias had no effect of exercise on  
25  
26 depression. Furthermore, duration of intervention and trial size were inversely associated with effect  
27  
28 estimates. Exercise did not improve quality of life or depression or remission after the intervention. Serious  
29  
30 adverse event or adverse events were reported inconsistently and only by a few trials not permitting firm  
31  
32 conclusions regarding these outcomes.  
33  
34  
35  
36  
37  
38  
39  
40

### 41 *Strengths and limitations*

42  
43 The strengths of this systematic review are that it is based on the published protocol, a comprehensive  
44  
45 search strategy, and the inclusion of patient centered outcomes such as quality of life as well as adverse  
46  
47 events. Also, to avoid spurious finding from repeated testing, Trial Sequential Analysis and Bayes analysis  
48  
49 were undertaken and these analyses did not suggest that the pooled estimates could be reduced to  
50  
51 random errors for effect on depression severity or no remission. Neither Trial Sequential Analysis nor Bayes  
52  
53 factor analysis are, however, able to wash of spurious effects induced by bias, or fraud or other  
54  
55 reasons.<sup>25,27;80-82</sup> Had we restricted the Trial Sequential Analysis to trials of potentially lower risk of bias, the  
56  
57  
58  
59  
60

1  
2  
3  
4 number of trials and participants would be limited and we had seen evidence far from crossing boundaries  
5  
6 for benefit, harms, or futility. The conclusions for serious adverse events and adverse events were  
7  
8 associated with wide confidence intervals due to lack of data and firm conclusions for these outcomes are  
9  
10 presently not available.  
11

12  
13  
14  
15  
16 The number of trials with adequate allocation concealment was 39% in the current systematic review  
17  
18 compared to only 15.1% in trials assessing non-drug interventions for depression.<sup>83</sup> Blinded outcome  
19  
20 assessment was performed in 52% of the included trials compared to 44% in non-drug antidepressant trials  
21  
22 in general.<sup>83</sup> The incomplete outcome bias domain was adequate in 48% of our included trials compared to  
23  
24 32.9% of antidepressant non-drug trials in general.<sup>83</sup> Compared to non-drug trials assessing interventions  
25  
26 for patients with depression, the included exercise trials have more bias domains with low risk of bias.  
27  
28 However, all our included trials were at high risk of bias. Two trials had low risk of bias for all bias domains  
29  
30 except for blinding of participants and trial personnel, and four trials fulfilled our criteria for trials at  
31  
32 potentially less risk of bias than the rest of the trials with at risk of bias. Despite a search strategy including  
33  
34 bibliographical databases and trials from China and South-America, the vast majority of included trials were  
35  
36 conducted in north America and western Europe, which is comparable to the geographical distribution of  
37  
38 non-drug trials in general<sup>83</sup> limiting the applicability to other geographic regions.  
39  
40  
41  
42  
43  
44

#### 45 *The effect of exercise on depression*

46  
47  
48 Our present results are similar to the latest Cochrane review by Cooney et al. (2013)<sup>23</sup> who found a  
49  
50 moderate effect of exercise on depressive symptoms (-0.62 SMD) when including all trials and no effect  
51  
52 when restricting the analysis to trials with less risk of bias (-0.18 SMD). The Cochrane review did find  
53  
54 evidence of a small antidepressant effect beyond the intervention, which we could not confirm in our  
55  
56 present systematic review. Bridle et al. (2012)<sup>12</sup> included 9 trials allocating old (> 60 years) patients with  
57  
58  
59  
60



1  
2  
3  
4 depression to exercise interventions versus control interventions. Restricting the analysis to four trials at  
5  
6 lower risk of bias they found small to moderate effect estimates (SMD -0.34) in favour of exercise. The  
7  
8 studies by Cooney et al.<sup>23</sup> and Bridle et al.<sup>12</sup> both included trials allocating patient with depressive  
9  
10 symptoms and not necessarily diagnosed using a validated diagnostic system, potentially explaining the  
11  
12 differences in the effect sizes. However, in our present systematic review the estimate for four trials at  
13  
14 potential less risk of bias than the remaining trials was -0.11 SMD and in the Cooney study the effect  
15  
16 estimate for eight trials with lower risk of bias was -0.18 SMD<sup>23</sup> compared to -0.34 in the study by Bridle at  
17  
18 al.<sup>12</sup> Meta-analysis of randomised clinical trials assessing the effects of exercise for depression consistently  
19  
20 finds positive effects, however, when restricting the analysis to trials with less risk of bias the pooled effect  
21  
22 sizes becomes very small or negligible. Meta-analysis examining the effect of exercise beyond the  
23  
24 intervention also finds no or small effects of exercise. In the process of interpretation of effect estimates in  
25  
26 the current research field, it is important to recognise that effect estimates from trials with non-blinded  
27  
28 outcome assessment are at high risk of bias as reported by Savovic et al.<sup>84</sup> Thirteen of 31 trials in the  
29  
30 current systematic review did not use blinded outcome assessment. In contradiction to the current  
31  
32 systematic review, a recent meta-analysis by Schuch et al.<sup>11</sup> concluded that “exercise has a large and  
33  
34 significant antidepressant effect in people with depression.....Our data strongly support the claim that  
35  
36 exercise is an evidence-based treatment for depression”. This statement was based on a meta-analysis of  
37  
38 25 randomised clinical trials including patients with depression or depressive symptoms to exercise or  
39  
40 control conditions and excluding trials using any form of active control group. Surprisingly, the authors  
41  
42 found that adjusting for publication bias using the Trim and Fill procedure<sup>29</sup> the estimate *increased* from a  
43  
44 SMD of 0.98 to 1.11. The effect in SMD in included studies ranged from -0.23 to 4.56 representing  
45  
46 considerable heterogeneity.<sup>11</sup> The authors classified four trials as having lower risk of bias using the same  
47  
48 criteria as in our systematic review and 21 trials as having high risk of bias. This illustrates some of the  
49  
50 challenges in meta-analysis of exercise and depression: the large heterogeneity driven by small studies  
51  
52 inflating the effects of random-effects analysis,<sup>85</sup> the misconception that we can restrict our analysis to  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 statistics and not consider the evident effect of bias.<sup>22;84</sup> We therefore recommend that future systematic  
5  
6 reviews and meta-analysis a priori should have a primary outcome restricting effect analysis to larger trials  
7  
8 with lower risk of bias and that any recommendations regarding exercise interventions for patients with  
9  
10 depression should be assessed with the GRADE framework.  
11

12  
13  
14  
15  
16 The  $I^2$  of 82% and 71% for the primary outcomes indicate substantial evidence of heterogeneity of  
17  
18 intervention effects that is variation in effect estimates beyond chance. Part of this heterogeneity was  
19  
20 explained by bias and by trial size: trials at high risk of bias or small trials have very large effect estimates  
21  
22 compared to trials potentially at less bias risk compared to the remaining trials at high risk of bias or larger  
23  
24 trials. The funnel plots and Egger's test indicates publication bias, however, the association between trial  
25  
26 size and effect estimates could suggest that the asymmetry in the funnel plots are due to small study bias  
27  
28 rather than publication bias.<sup>86</sup> In addition, in line with our previous review we found duration of  
29  
30 intervention inversely associated with effect size.<sup>10</sup> A number of studies compare exercise to control  
31  
32 interventions rather than wait-list control to reduce the effect of non-specific effects, e.g., the DEMO trials  
33  
34 and Mather et al.<sup>20;21;50</sup> Also, it could be speculated that the effect of exercise would be harder to detect if  
35  
36 patients also received medical treatment in addition. The current systematic review could not confirm that  
37  
38 the type of control condition explained heterogeneity. The discussion of control group is important in non-  
39  
40 drug trials: choosing a waitlist control group the results potentially reflects non-specific effects, choosing an  
41  
42 active control group (e.g., relaxation exercise) the trial is potentially a comparison between to active  
43  
44 treatments. However, in the current systematic review we found no evidence that trials using an attention  
45  
46 control group or exercise as add-on to pharmacotherapy had significantly different effect estimates  
47  
48 compared to other trials.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 Our systematic review did not find indications of a positive effect on quality of life in patients with  
5 depression allocated to exercise interventions, which is in concordance with the review by Cooney et al.<sup>23</sup>  
6  
7  
8 Only 3/31 trials reported on serious adverse events, and found no significant risk of death or suicide  
9  
10 attempt. No indication of increased severity of depression or other adverse events in participants allocated  
11 to exercise could be detected. However, data on adverse events was reported sporadically in a minority of  
12 trials and currently it is not possible to conclude on the risk of serious adverse events or adverse event from  
13 exercise interventions in patients with depression.  
14  
15  
16  
17  
18  
19  
20  
21

### 22 *Conclusions*

23  
24 We have little confidence in the pooled effect estimates, especially because trials with less than high risk of  
25 bias produced significantly lower effect estimates, suggesting that exercise interventions only produce  
26 small or negligible antidepressant effects, depending on how much of the effect is caused by bias and how  
27 much is caused by the intervention. There was no effect of exercise on quality of life or depression beyond  
28 the intervention itself. There is currently no evidence in favour of exercise for patients with depression with  
29 a view to ameliorate depressive symptoms and at we do not recommend that exercise is prescribed to  
30 relieve depressive symptoms. Our systematic review did not evaluate possible beneficial effects of exercise  
31 on, e.g., metabolism or cardiovascular fitness,<sup>21;87</sup> and it is possible that exercise may have beneficial effects  
32 on these factors in patients diagnosed with depression.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

### 49 *Future perspectives*

50  
51 Despite the large number of published trials, further trials with more robust methodology seem still  
52 required to establish progress in this field. Also, additional trials from outside North-America and Europe  
53 may be required for results to be valid for patients in Asia, Africa, and South-America. To further elaborate  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 on the current findings, we recommend that future trials must include blinded outcome assessors and  
5  
6 outcomes assessing quality of life, metabolic effects, and long-term effects beyond the intervention. It is  
7  
8 also important that future trials systematically collect and report data on death, suicide events,  
9  
10 musculoskeletal injuries and other potential adverse effects in both the intervention group as well as in the  
11  
12 control group. Moreover, future trials ought to be designed according to the SPIRIT guidelines and reported  
13  
14 according to the CONSORT guidelines<sup>88,89</sup> and transparently report deidentified individual patient data  
15  
16 enabling individual patient data meta-analyses.<sup>90</sup>  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Funding

No funding was obtained for this systematic review.

## Competing interests

JK, CG, and MN have previously published two trials and a meta-analysis on this topic, which could introduce an academic bias in the current systematic review. We asked new authors to be involved in the preparation of the protocol, trial selection and bias assessment. No support from any organisation was received for the submitted work; no financial relationship with any organisations that might have an interest in the submitted work in the previous three years, and apart from the above no other relationship or activities that could appear to have influenced the submitted work.

## Acknowledgement

We appreciate the help from Youling He for help with the Chinese Wanfang bibliographical database and translation of Chinese papers. We thank Janus C Jakobsen for assistance with the calculation of Bayes factor.

## Contributors

JK conceived the project, collected data, did the statistical analysis, analysed the data, drafted and revised the manuscript. He is guarantor. CH collected the data, analysed the data and revised the manuscript. HS conceived the project, collected data, analysed the data, and revised the manuscript. CG conceived the project, analysed the data and revised the manuscript. MN conceived the project, analysed the data, and revised the manuscript.

## References

- 1  
2  
3  
4  
5  
6  
7  
8 (1) Lepine JP, Gastpar M, Mendelwicz J, Tylee A. Depression in the community: The first pan-  
9 european study DEPRES (Depression Research in European Society). *Int Clin Psychopharmacol*  
10 1997; 12:19-29.
- 11  
12 (2) Pedersen CB, Mors O, Bertelsen A, Waltoft BL, et al. A comprehensive nationwide study of the  
13 incidence rate and lifetime risk for treated mental disorders. *JAMA Psychiatry* 2014; 71(5):573-  
14 581.
- 15  
16 (3) Ustun TB, Ayuso-Mateos JL, Chatterji S, Mathers C, Murray CJL. Global burden of depressive  
17 disorders in the year 2000. *Br J Psychiatry* 2004; 184:386-392.
- 18  
19 (4) Kirsch I, Deacon B, Huedo-Medina T, Scoboria A, Moore T, Johnson B. Initial severity and  
20 antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration.  
21 *PLoS Med* 2008; 5[2], e45. doi:10.1371/journal.pmed.0050045.
- 22  
23 (5) Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective Publication of  
24 Antidepressant Trials and Its Influence on Apparent Efficacy. *N Engl J Med* 2008; 358(3):252-260.
- 25  
26 (6) Jakobsen JC, Hansen JL, Storebo OJ, Simonsen E, Gluud C. The effects of cognitive therapy versus  
27 'no intervention' for major depressive disorder. *PLoS One* 2011; 6(12):e28299.
- 28  
29 (7) Jakobsen JC, Hansen JL, Simonsen E, Gluud C. The effect of interpersonal psychotherapy and  
30 other psychodynamic therapies versus 'treatment as usual' in patients with major depressive  
31 disorder. *PLoS One* 2011; 6(4):e19044.
- 32  
33 (8) Jakobsen JC, Hansen JL, Simonsen E, Gluud C. The effect of adding psychodynamic therapy to  
34 antidepressants in patients with major depressive disorder. A systematic review of randomized  
35 clinical trials with meta-analyses and trial sequential analyses. *J Affect Disord* 2012; 137(1-3):4-14.
- 36  
37 (9) Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R et al. Comparative efficacy  
38 and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis.  
39 *Lancet* 2009; 373[9665]: 746-758.
- 40  
41 (10) Krogh J, Nordentoft M, Sterne J, Lawlor D. The effect of exercise in clinically depressed adults:  
42 Systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry* 2011;  
43 72(4): 529-538.
- 44  
45 (11) Schuch FB, Vancampfort D, Richards J, Rosenbaum S, Ward PB, Stubbs B. Exercise as a treatment  
46 for depression: A meta-analysis adjusting for publication bias. *J Psychiatr Res* 2016; 77:42-51.
- 47  
48 (12) Bridle C, Spanjers K, Patel S, Atherton NM, Lamb SE. Effect of exercise on depression severity in  
49 older people: systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry*  
50 2012; 201(3):180-185.
- 51  
52 (13) Herring MP, Puetz TW, O'Connor PJ, Dishman RK. Effect of exercise training on depressive  
53 symptoms among patients with a chronic illness: a systematic review and meta-analysis of  
54 randomized controlled trials. *Arch Intern Med* 2012; 172(2):101-111.
- 55  
56  
57  
58  
59  
60

- 1
- 2
- 3
- 4 (14) World Health Organisation. International Statistical Classification of Diseases, 10th Revision (ICD-
- 5 10). Geneva, Switzerland, 1992.
- 6
- 7 (15) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed.
- 8 Washington DC, 1994.
- 9
- 10 (16) Beck AT, Steer RA, Brown GK. BDI-II manual. 2 ed. New York: Psychological Corporation, 1996.
- 11
- 12
- 13 (17) Krogh J, Speyer H, Gluud C, Nordentoft M. Exercise for patients with major depression: A
- 14 systematic review with meta-analysis and trial sequential analysis. *Syst Rev* 2015; 4(1):40.
- 15
- 16 (18) Spitzer RL, Endicott J, Williams JB. Research diagnostic criteria. *Arch Gen Psychiatry* 1979;
- 17 36(12):1381-1383.
- 18
- 19 (19) ICH-secretariat. International Conference on Harmonisation (ICH) Harmonised Tripartite
- 20 Guideline (Step 4). Guideline for Good Clinical Practice. Geneva, 1996.
- 21
- 22
- 23 (20) Krogh J, Saltin B, Gluud C, Nordentoft M. The DEMO trial: A randomized, parallel-group,
- 24 observer-blinded clinical trial of strength versus aerobic versus relaxation training for patients
- 25 with mild to moderate depression. *J Clin Psychiatry* 2009; 70[6]: 790-800.
- 26
- 27
- 28 (21) Krogh J, Videbech P, Thomsen C, Gluud C, Nordentoft M. DEMO-II trial. Aerobic Exercise Versus
- 29 Stretching Exercise in Patients With Major Depression - A Randomised Clinical Trial. *PLoS One*
- 30 2012; 7(10):e48316. doi: 10.1371/journal.pone.0048316.
- 31
- 32
- 33 (22) Higgins JPT. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The
- 34 Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org) The Cochrane
- 35 Collaboration, 2009.
- 36
- 37
- 38 (23) Cooney GM, Dwan K, Greig CA, Lawlor DA, Rimer J, Waugh FR et al. Exercise for depression.
- 39 *Cochrane Database Syst Rev* 2013; 9:CD004366.
- 40
- 41 (24) Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ*
- 42 2003; 327:557-560.
- 43
- 44 (25) Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm
- 45 evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008; 61(1):64-75.
- 46
- 47 (26) Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information
- 48 size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008; 61(8):763-
- 49 769.
- 50
- 51 (27) Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical
- 52 significance in systematic reviews with meta-analytic methods. *BMC Med Res Methodol* 2014;
- 53 14:120.
- 54
- 55
- 56
- 57
- 58
- 59
- 60



- 1  
2  
3  
4 (28) Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published  
5 randomised controlled trials. *BMJ* 1999; 319(7211):670-674.  
6  
7 (29) Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for  
8 publication bias in meta-analysis. *Biometrics* 2000; 56(2):455-463.  
9  
10 (30) Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P et al. GRADE: an emerging  
11 consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;  
12 336(7650):924-926.  
13  
14 (31) Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y et al. GRADE guidelines: 14.  
15 Going from evidence to recommendations: the significance and presentation of  
16 recommendations. *J Clin Epidemiol* 2013; 66(7):719-725.  
17  
18 (32) Zheng YP, Lin KM, Zhao JP, Zhang MY, Yong D. Comparative study of diagnostic systems: Chinese  
19 Classification of Mental Disorders-Second Edition versus DSM-III-R. *Compr Psychiatry* 1994;  
20 35(6):441-449.  
21  
22 (33) Babyak MA, Blumenthal JA, Herman S, Khatri P, Doraiswamy M, Moore K. Exercise treatment for  
23 major depression: Maintenance of therapeutic benefit at 10 months. *Psychosom Med* 2000;  
24 62:633-638.  
25  
26 (34) Baxter H, Winder R, Chalder M, Wright C, Sherlock S, Haase A et al. Physical activity as a  
27 treatment for depression: the TREAD randomised trial protocol. *Trials* 2010; 11:105.  
28  
29 (35) Benson M, Blumenthal J, Babyak M, Smith P, Rogers S, Doraiswamy M et al. Exercise fails to  
30 improve neurocognition in depressed middle-aged and older adults. *Med Sci Sports Exerc* 2008;  
31 40[7]: 1344-1352.  
32  
33 (36) Blumenthal J, Babyak MA, Moore KA, Craighead WE, Herman S, Khatri P. Effects of exercise  
34 training on older patients with major depression. *Arch Intern Med* 1999; 159:2349-2356.  
35  
36 (37) Blumenthal J, Babyak M, Doraiswamy P, Watkins L, Hoffman B, Barbour K et al. Exercise and  
37 pharmacotherapy in the treatment of major depressive disorder. *Psychosom Med* 2007; 69: 587-  
38 596.  
39  
40 (38) Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Haase AM, Taylor AH et al. Facilitated physical  
41 activity as a treatment for depressed adults: randomised controlled trial. *BMJ* 2012; 344:e2758.  
42  
43 (39) Chalder M, Wiles NJ, Campbell J, Hollinghurst SP, Searle A, Haase AM et al. A pragmatic  
44 randomised controlled trial to evaluate the cost-effectiveness of a physical activity intervention  
45 as a treatment for depression: the treating depression with physical activity (TREAD) trial. *Health*  
46 *Technol Assess* 2012; 16(10):1-iv.  
47  
48 (40) Combs K, Smith PJ, Sherwood A, Hoffman B, Carney RM, Freedland K et al. Impact of sleep  
49 complaints and depression outcomes among participants in the standard medical intervention  
50 and long-term exercise study of exercise and pharmacotherapy for depression. *J Nerv Ment Dis*  
51 2014; 202: 167-171.  
52  
53 (41) Doyne EJ, Ossip-Klein DJ, Bowman ED, Osborn KM, McDougall-Wilson IB, Neimeyer RA. Running  
54 versus weight lifting in the treatment of depression. *J Consult Clin Psychol* 1987; 55(5):748-754.  
55  
56  
57  
58  
59  
60

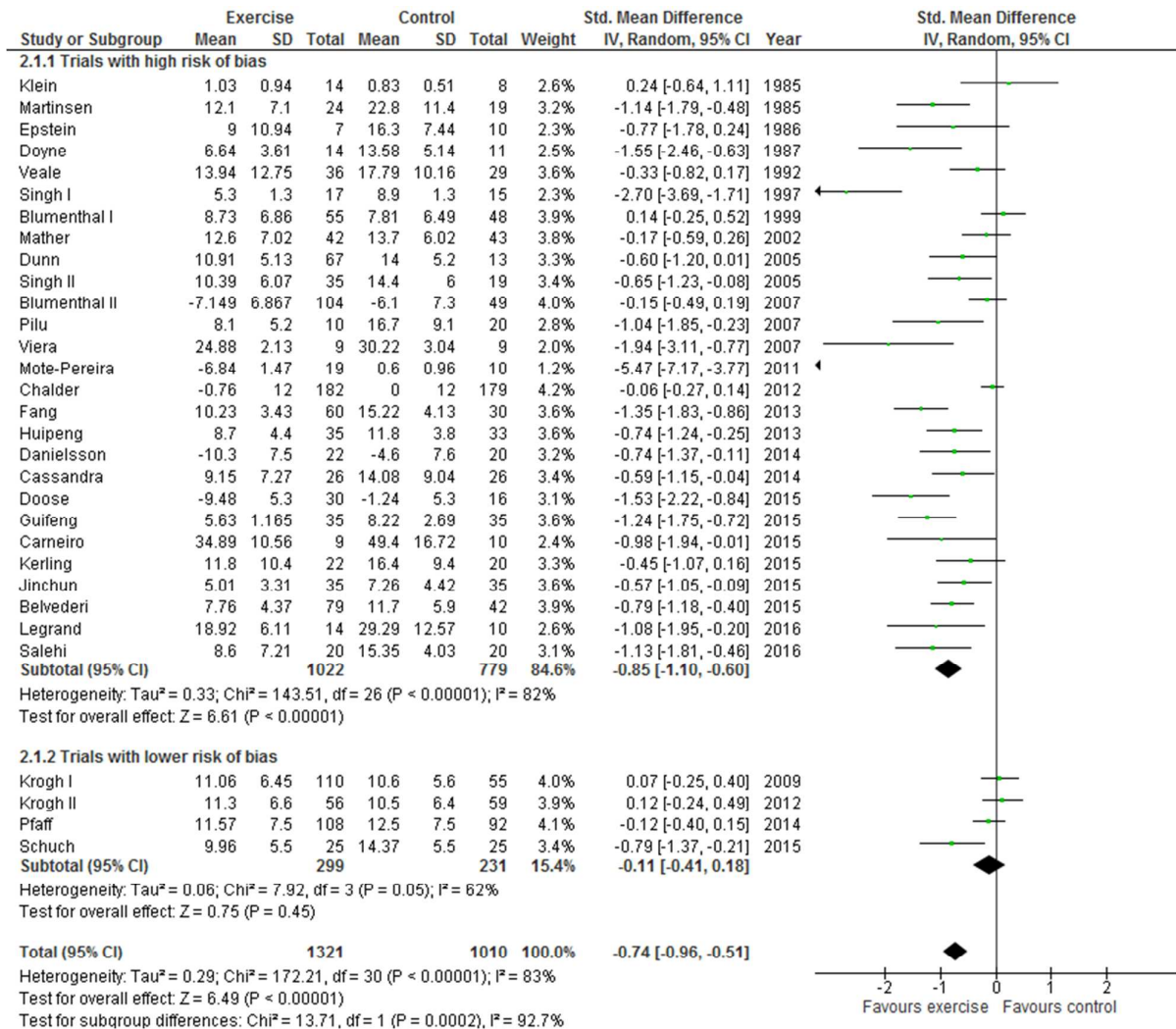
- 1  
2  
3  
4 (42) Dunn AL, Trivedi MH, Kampert JB, Clark CG, Chambliss HO. The DOSE study: a clinical trial to  
5 examine efficacy and dose response of exercise as treatment for depression. *Control Clin Trials*  
6 2002; 23(5):584-603.  
7  
8 (43) Dunn AL, Trivedi MH, Kaplan K, Clark C, Chambliss H. Exercise treatment for depression. Efficacy  
9 and dose response. *Am J Prev Med* 2005; 28: 1-8.  
10  
11 (44) Herman S, Blumenthal JA, Babyak M, Khatri P, Craighead WE, Krishnan KR et al. Exercise therapy  
12 for depression in middle-aged and older adults: predictors of early dropout and treatment failure.  
13 *Health Psychol* 2002; 21:553-563.  
14  
15 (45) Ho CWH, Chan SC, Wong JS, Cheung WT, Chung DWS, Lau TFO. Effect of aerobic exercise training  
16 on chinese population with mild to moderate depression in Hong Kong. *Rehabil Res Pract*  
17 2014;2014:627376. doi: 10.1155/2014/627376  
18  
19 (46) Hoffman BM, Babyak MA, Craighead WE, Sherwood A, Doraiswamy PM, Coons MJ et al. Exercise  
20 and pharmacotherapy in patients with major depression: one-year follow-up of the SMILE study.  
21 *Psychosom Med* 2011; 73(2):127-133.  
22  
23 (47) Kerling A, Tegtbur U, Gutzlaff E, Kuck M, Borchert L, Ates Z et al. Effects of adjunctive exercise on  
24 physiological and psychological parameters in depression: A randomized pilot trial. *J Affect Disord*  
25 2015: 177:1-6.  
26  
27 (48) Khatri P, Blumenthal J, Babyak M, Craighead E, Herman S, Baldewicz T et al. Effects of exercise  
28 training on cognitive functioning among depressed older men and women. *J Aging Phys Act* 2001;  
29 9: 43-57.  
30  
31 (49) Krogh J, Petersen L, Timmermann M, Saltin B, Nordentoft M. Design paper: The DEMO trial: A  
32 randomized, parallel-group, observer-blinded clinical trail of aerobic versus non-aerobic versus  
33 relaxation training for patients with light to moderate depression. *Contem Clin Trials* 2007; 28:79-  
34 89.  
35  
36 (50) Mather AS, Rodriguez C, Guthrie MF, McHarg AM, Reid IC, McMurdo ME. Effects of exercise on  
37 depressive symptoms in older adults with poorly responsive depressive disorder: randomised  
38 controlled trial. *Br J Psychiatry* 2002; 180:411-415.  
39  
40 (51) Mota-Pereira J, Silverio J, Carvalho S, Ribeiro JC, Fonte D, Ramos J. Moderate exercise improves  
41 depression parameters in treatment-resistant patients with major depressive disorder. *J Psychiatr*  
42 *Res* 2011; 45(8):1005-1011.  
43  
44 (52) Pfaff JJ, Alfonso H, Newton RU, Sim M, Flicker L, Almeida OP. ACTIVEDEP: a randomised,  
45 controlled trial of a home-based exercise intervention to alleviate depression in middle-aged and  
46 older adults. *Br J Sports Med* 2014; 48(3):226-232.  
47  
48 (53) Pilu A, Sorba M, Hardoy M, Floris A, Mannu F, Seruis M et al. Efficacy of physical activity in the  
49 adjunctive treatment of major depressive disorders: preliminary results. *Clin Pract Epidemiol*  
50 *Ment Health* 2007; 3(8).  
51  
52 (54) Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Fleck MP. Exercise and severe depression:  
53 preliminary results of an add-on study. *J Affect Disord* 2011; 133(3):615-618.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (55) Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of the effect of exercise on  
5 sleep. *Sleep* 1997; 20(2):95-101.  
6  
7 (56) Singh NA, Stavrinou TM, Scarbek Y, Galambos G, Liber C, Fiatarone Singh MA. A randomized  
8 controlled trial of high versus low intensity weight training versus general practitioner care for  
9 clinical depression in older adults. *J Gerontol A Biol Sci Med Sci* 2005; 60(6):768-776.  
10  
11 (57) Singh NA, Clements KM, Singh MA. The efficacy of exercise as a long-term antidepressant in  
12 elderly subjects: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* 2001; 56(8):M497-  
13 M504.  
14  
15 (58) Singh N, Clements K, Fiatarone M. A randomized controlled trial of progressive resistance training  
16 in depressed elders. *J Gerontol* 1997; 52:M27-M35.  
17  
18 (59) Toni G, Belvederi MM, Zanetidou S, Menchetti M, Ermini G, Tripi F et al. Physical exercise for late  
19 life Major Depression: The SEEDS study. *Eur J Prev Cardiol* 2014; 21:S107.  
20  
21 (60) Veale D, Le FK, Pantelis C, de S, V, Mann A, Sargeant A. Aerobic exercise in the adjunctive  
22 treatment of depression: a randomized controlled trial. *J R Soc Med* 1992; 85(9):541-544.  
23  
24 (61) Vieira JLL, Porcu M, da Rocha PGM. The regular physical exercise's practice as complementary  
25 therapy to the treatment of depressed women. *J Bras Psiquiatr* 2007; 56: 23-28.  
26  
27 (62) Epstein D. Reno, Dissertation: University of Nevada:Aerobic activity versus group cognitive  
28 therapy: An evaluative study of contrasting interventions for the alleviation of clinical depression.  
29 1986.  
30  
31 (63) Danielsson L, Papoulias I, Petersson E, Carlsson J, Waern M. Exercise or basic body awareness  
32 therapy as add-on treatment for major depression: A controlled study. *J Affect Disord* 2014;  
33 168:98-106.  
34  
35 (64) Carneiro L, Fonseca A, Vieira-Coelho M, Mota M, Vasconcelas-Raposo J. Effects of structured  
36 exercise and pharmacotherapy vs pharmacotherapy for adults with depressive symptoms: A  
37 randomized clinical trial. *J Psych Res* 2015;(71):48-55.  
38  
39 (65) Legrand FD, Neff EM. Efficacy of exercise as an adjunct treatment for clinically depressed  
40 inpatients during the initial stages of antidepressant pharmacotherapy: An open randomized  
41 controlled trial. *J Affect Disord* 2016; 191:139-144.  
42  
43 (66) Doose M, Ziegenbein M, Hoos O, Reim D, Stengert W, Hoffer N et al. Self-selected intensity  
44 exercise in the treatment of major depression: A pragmatic RCT. *Int J Psychiatry Clin Pract* 2015;  
45 19(4):266-275.  
46  
47 (67) Salehi I, Hosseini SM, Haghghi M, Jahangard L, Bajoghli H, Gerber M et al. Electroconvulsive  
48 therapy (ECT) and aerobic exercise training (AET) increased plasma BDNF and ameliorated  
49 depressive symptoms in patients suffering from major depressive disorder. *J Psychiatr Res* 2016;  
50 76:1-8.  
51  
52 (68) Huipeng F, Xiaohui Z. Curative effect and observation of exercise on patients with depression.  
53 *Chin Med Pharm* 2013; 3:187-188.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (69) Jinchun W, Bo Y, Rong X, Chengfeng L, Yanan S. Effect of exercise therapy in improving social  
5 function and quality of life in patients with depression. *Nurs J Chin PLA* 2015; 32(21):24.  
6  
7 (70) Guifeng S. The role of setting - up exercise on rehabilitation in patients with depressive episode.  
8 *Chin J Health Psychol* 2015; 23:184-186.  
9  
10 (71) Fang D, Min R, Ai-qin W. The effect of aerobic exercise with different intensity and frequency on  
11 depressive symptoms. *Chin J Behav Med & Brain Sci* 2013; 22:830-832.  
12  
13 (72) Martinsen EW, Medhus A, Sandvik L. Effects of aerobic exercise on depression: A controlled  
14 study. *BMJ* 1985; 291:109.  
15  
16 (73) Veale D, Fevre K, Pantalís C, Souza V, Mann A, Sargeant A. Aerobic exercise in the adjunctive  
17 treatment of depression: A randomised controlled trial. *J R Soc Med* 1992; 85:541-544.  
18  
19 (74) Ossip-Klein DJ, Doyne EJ, Bowman ED, Osborn KM, Dougall-Wilson IB, Neimeyer RA. Effects of  
20 running or weight lifting on self-concept in clinically depressed women. *J Consult Clin Psychol*  
21 1989; 57(1):158-161.  
22  
23 (75) Blumenthal JA, Babyak MA, Moore KA, Craighead WE, Herman S, Khatri P et al. Effects of exercise  
24 training on older patients with major depression. *Arch Intern Med* 1999; 159(19):2349-2356.  
25  
26 (76) Belvederi MM, Amore M, Menchetti M, Toni G, Neviani F, Cerri M et al. Physical exercise for late-  
27 life major depression. *Br J Psychiatry* 2015; 207(3):235-242.  
28  
29 (77) Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Zimmermann AB, Rocha NS, Fleck MP. Exercise  
30 and severe major depression: effect on symptom severity and quality of life at discharge in an  
31 inpatient cohort. *J Psychiatr Res* 2015; 61:25-32.  
32  
33 (78) Carta MG, Hardoy MC, Pílu A, Sorba M, Floris AL, Mannu FA et al. Improving physical quality of  
34 life with group physical activity in the adjunctive treatment of major depressive disorder. *Clin*  
35 *Pract Epidemiol Ment Health* 2008; 4:1  
36  
37 (79) Klein MH, Greist JH, Gurman AS, Neimeyer R, Lesser D, Bushnell N et al. A comparative outcome  
38 study of group psychotherapy vs. exercise treatments for depression. *Int J Ment Health* 1985;  
39 13:148-177.  
40  
41 (80) Roberts I, Ker K, Edwards P, Beecher D, Manno D, Sydenham E. The knowledge system  
42 underpinning healthcare is not fit for purpose and must change. *BMJ* 2015; 350:h2463.  
43  
44 (81) Gluud C, Jakobsen J, Imberger G, Lange T, Wetterslev J. Re: The knowledge system underpinning  
45 healthcare is not fit for purpose and must change - Responses to the opposing viewpoints of  
46 Roberts and colleagues and Tovey and colleagues. *BMJ* 2015; 350(h2463).  
47  
48 (82) Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for Trial  
49 Sequential Analysis (TSA). Copenhagen Trial Unit ,Centre for Clinical Intervention Research,  
50 Copenhagen, Available from <https://www.ctu.dk/tsa>, 2011.  
51  
52 (83) Krogh J, Hjorthoj CR, Jakobsen JC, Lindschou J, Kessing LV, Nordentoft M et al. DEPEROR: Risks  
53 of systematic errors in drug and non-drug randomized clinical trials assessing intervention effects  
54 in patients with unipolar depression. *J Affect Disord* 2015; 179:121-127.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (84) Savovic J, Jones HE, Altman DG, Harris RJ, Juni P, Pildal J et al. Influence of reported study design  
5 characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern*  
6 *Med* 2012; 157(6):429-438.  
7  
8 (85) Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol*  
9 1999; 150(5):469-475.  
10  
11 (86) Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J et al. Recommendations for examining  
12 and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*  
13 2011; 343:d4002.  
14  
15 (87) Stubbs B, Rosenbaum S, Vancampfort D, Ward PB, Schuch FB. Exercise improves  
16 cardiorespiratory fitness in people with depression: A meta-analysis of randomized control trials.  
17 *J Affect Disord* 2016; 190:249-253.  
18  
19 (88) Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting  
20 parallel group randomized trials. *Ann Intern Med* 2010; 152(11):726-732.  
21  
22 (89) Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA et al. SPIRIT 2013 explanation  
23 and elaboration: guidance for protocols of clinical trials. *BMJ* 2013; 346:e7586.  
24  
25 (90) Skoog M, Saarimäki J, Gluud C, Sheinin M, Erlandsson K, Aamdal S. Transparency and registration  
26 in clinical research in the Nordic countries. Nordic Trial Alliance, NordForsk; 2015: 1-108.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Figure 1. Effect of exercise on depression severity in patients diagnosed with depression



BMJ Open: first published as 10.1136/bmjopen-2016-014820 on 18 September 2017. Downloaded from http://bmjopen.bmj.com/ on April 18, 2024 by guest. Protected by copyright.

**Table 1.** Characteristics of trials assessing exercise for patients diagnosed with depression

Author, first Country of origin	Participants	Severity of depression at baseline	N at baseline (included in trial efficacy analysis)	Type of intervention	Frequency	Duration
Klein 1985 USA	Outpatients Mean age: 30 (SD 7) 72% female	SCL-D: 2.4 (SD 1)	50 (22)	<i>Aerobic exercise:</i> Supervised individual running. <i>Control group:</i> Supervised meditation in groups	2 sessions per week  Control group: 1 session per week	12 weeks
Martinsen 1985 Norway	Inpatients Mean age: 40 (range 17-60) Distribution of sex not reported	BDI: 28.0 (SD 9)	49(43)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Occupational therapy.	3 sessions per week  Control group: 3 sessions per week	9 weeks
Epstein 1986 USA	Outpatients Mean age: 39 (range 24 to 60) (NR) % female	BDI: 23.4 (SD 7)	21 (17)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Waitlist control.	3 sessions per week	8 weeks
Doyne 1987 USA	Outpatients Mean age: 29 (SD 4) 100 % female	HAM-D <sub>17</sub> : 13.0 (SD 7)	52 (25)	<i>Aerobic exercise OR weightlifting:</i> Supervised individual exercise. <i>Control group:</i> Waiting list.	4 sessions per week	8 weeks
Veale 1992 UK	Outpatients Mean age: 35 (range 19-58) 64% female	BDI: 24.5 (SD 6)	83 (65)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Standard treatment from psychiatric services.	3 sessions per week	12 weeks
Singh 1997 USA	Outpatients Recruited from a register of volunteers Mean age: 71 (SD 1)	BDI: 19.9 (SD 2.3)	32 (32)	<i>Progressive resistance training:</i> Supervised group exercise. <i>Control group:</i> Attended seminars on health.	3 sessions per week  Control group: 2 sessions per week	10 weeks
Blumenthal 1999 USA	Outpatients Mean age: 57 (SD 7) 71.8% female	HAM-D <sub>17</sub> : Not reported	103 (103)	<i>Aerobic exercise:</i> Supervised exercise plus antidepressant medication (sertraline). <i>Control group:</i> Antidepressant medication (sertraline).	3 sessions per week	16 weeks
Mather 2002 UK	Outpatients Treatment resistant Mean age: 65 (range 53-91) 69% female	HAM-D <sub>17</sub> : 17.1 (SD 6)	86 (85)	<i>Mixed aerobic and non-aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Attended health seminars.	2 sessions per week  Control group: 2 seminars per week	10 weeks
Dunn 2005 USA	Outpatients Mean age: 36 (SD 6)	HAM-D <sub>17</sub> : 19.4 (SD 2)	80 (80)	<i>Aerobic exercise:</i> Individually supervised	Group (1) and (2): 3 sessions	12 weeks

		75% female			exercise with (1) low energy expenditure (EE) OR (2) high EE OR (3) low EE OR (4) high EE. <i>Control group:</i> Flexibility exercise.	per week Group (3) and (4): 5 sessions per week Control group: 3 sessions per week	
Singh 2005 Australia	Outpatients Mean age: 69 (SD 6) 55% female	HAM-D <sub>17</sub> : 18.9 (SD 4.2)	60 (54)	<i>Progressive resistance training (PRT):</i> (1) Low intensity PRT OR (2) high intensity PRT. <i>Control group:</i> Standard GP care.	Group (1) and (2): 3 sessions per week	8 weeks	
Pilu 2007 Italy	Outpatients Treatment resistant Age between 40 and 60 100% female	HAM-D <sub>17</sub> : 19.7 (SD 6)	30 (30)	<i>Resistance exercise:</i> Supervised group sessions. <i>Control group:</i> Standard treatment.	2 sessions per week	32 weeks	
Viera 2007 Brazil	Outpatients Mean age 43.66 (SD NR) 100% female	HAM-D <sub>21</sub> : 31.9 (SD 3)	18 (18)	<i>Aerobic exercise:</i> Supervised water aerobics. <i>Control group:</i> Standard GP care.	2 sessions per week	12 weeks	
Blumenthal 2007 USA	Outpatients Mean age: 52 (SD 8) 75.8% female	HAM-D <sub>17</sub> : 16.7 (SD 4)	153 (153)	<i>Aerobic exercise:</i> (1) Supervised group exercise OR (2) home-based exercise. <i>Control group:</i> Placebo medication.	(1) and (2): 3 sessions per week	16 weeks	
Krogh 2009 Denmark	Outpatients Mean age: 39 (SD 9) 74% female	HAM-D <sub>17</sub> : 17.8 (SD 4)	165 (165)	<i>Exercise:</i> (1) Aerobic supervised group exercise OR (2) supervised group resistance training <i>Control group:</i> relaxation and stretching exercise.	(1) and (2): 2 sessions per week  Control group: 2 sessions per week	16 weeks	
Mota-Pereira 2011 Portugal	Outpatients Treatment resistant Mean age: 47.5 (SD 3) 65.5% female	HAM-D <sub>17</sub> : 17.1 (SD 3)	33 (29)	<i>Aerobic exercise:</i> Homebased exercise + supervised. <i>Control group:</i> Attention control.	4 home-based sessions/week. 1 supervised session/week Control group: 1 supervised session/week	12 weeks	
Krogh 2012 Denmark	Outpatients Mean age: 42 (SD 11) 67% female	HAM-D <sub>17</sub> : 18.9 (SD 4)	115 (115)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Supervised stretching exercise in groups.	3 sessions per week  Control group: 3 sessions per week	12 weeks	
Chalder 2012 UK	Outpatients Mean age: 40 (SD 13) 66% female	BDI: 32.1 (SD 9)	361 (361)	<i>Exercise:</i> Participants received individually tailored support and encouragement to engage in physical activity. <i>Control group:</i> Standard GP care.	Individual	16 weeks	



1						
2						
3						
4	Fang 2013	Inpatients	HAM-D <sub>24</sub> :	90 (90)	<i>Aerobic exercise:</i>	Group 1 and 2
5	China	Mean age: 44 (SD	29.2 (SD 5)		Group 1 and 2 had 3 and 5	6 weeks
6		14)			supervised group	
7		66.9% female			exercise, high	sessions per
8					intensity.	week,
9					<i>Control group:</i>	respectively
10					15 min stretching	Control group:
11	Huipeng 2013	Inpatients	HAM-D <sub>17</sub> :	68 (68)	<i>Aerobic exercise:</i>	5 sessions per
12	China	Mean age: 30 (SD 5)	28 (SD 5)		Jogging	week
13		100% female			<i>Control group:</i>	
14					Standard treatment	
15	Cassandra 2014	Inpatients	MADRS:	52 (52)	<i>Aerobic exercise:</i>	5 sessions per
16	Honkong	Mean age: 46 (SD	19 (10)		Supervised exercise.	week
17		12)			<i>Control group:</i>	
18		67.3% female			10 min stretching.	3 weeks
19	Danielsson 2014	Outpatients	MADRS:	42 (42)	<i>Mixed aerobic and</i>	2 sessions per
20	Sweden	Mean age: 45 (SD	24.0 (SD 5)		<i>non-aerobic exercise:</i>	week
21		13)			First two weeks	
22		76% female			individual supervised	
23					exercise then	
24					supervised group	
25					exercise.	
26					<i>Control group:</i> One	
27					session with advice on	
28	Pfaff 2014	Outpatients	MADRS:	200 (200)	<i>Resistance exercise:</i>	3 sessions per
29	Australia	Mean age: 61 (SD 8)	21.3 (SD		Supervised home-	week
30		63% female	NR)		based exercise	12 weeks
31					<i>Control group:</i>	
32					Standard GP care	
33	Guifeng 2015	Inpatients	HAM-D <sub>24</sub> :	70 (70)	<i>Aerobic exercise:</i>	5 sessions per
34	China	Mean age: 33 (SD	25.9 (SD 4)		Supervised group	week
35		14)			exercise	8 weeks
36		70% female			<i>Control group:</i>	
37					Standard treatment	
38	Junchin 2015	Inpatients	HAM-D <sub>24</sub> :	70 (70)	<i>Aerobic exercise:</i>	5 sessions per
39	China	Mean age: 28 (SD 7)	25.8 (SD 3)		Supervised aerobic	week
40		61% female			exercise of the	8 weeks
41					patients own choice	
42					<i>Control group:</i>	
43					Standard treatment	
44	Schuch 2015	Inpatients	HAM-D <sub>17</sub> :	50 (50)	<i>Aerobic exercise:</i>	3 sessions per
45	Brazil	Mean age: 40 (SD	26.7 (SD 2)		Supervised individual	week
46		11)			exercise.	2 weeks
47		74% female			<i>Control group:</i>	
48					Standard treatment.	
49	Kerling 2015	Inpatients	MADRS:	42 (42)	<i>Aerobic exercise:</i>	3 sessions per
50	Germany	Mean age: 43 (SD	24.0 (SD 9)		Supervised exercise.	week
51		10)			<i>Control group:</i>	
52					Standard treatment.	
53	Belvederi 2015	Outpatients	HAM-D <sub>17</sub> :	121 (121)	<i>Aerobic exercise:</i>	3 sessions per
54	Italy	Mean age: 75 (SD 6)	20.1 (SD 3)		(1) Sertraline +	week
55		71% female			supervised non-	24 weeks
56					progressive exercise	
57					OR (2) sertraline +	
58					supervised	
59					progressive aerobic	
60						

				exercise. <i>Control group:</i> Sertraline.		
Carneiro 2015 Portugal	Outpatients Mean age: 50.16 (SD 12) 100% female	BDI: 48.8 (SD 10)	26 (19)	<i>Aerobic exercise:</i> Supervised exercise <i>Control group:</i> Standard treatment	3 sessions per week	16 weeks
Doose 2015 Germany	Outpatients Mean age: 47.9 (SD 10.5) 63% female	HAM-D <sub>17</sub> : 14.2 (SD 3)	46 (46)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Standard treatment	3 sessions per week	8 weeks
Salehi 2016 Iran	Inpatients Mean age: 30.0 (SD 6) 35% female	HAM-D <sub>21</sub> : 43.4 (SD 8)	40 (40)	<i>Aerobic exercise + ECT:</i> Supervised aerobic exercise <i>Control group:</i> ECT	3 sessions per weeks  Control group 3 ECTs per week	4 weeks
Legrand 2016 France	Inpatients Mean age: 46.9 (SD 13) 67% female	BDI: 36.0 (SD 6)	24 (24)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Standard treatment	10 sessions in 10 consecutive days	10 days

SCL-D: Symptom Check List, depression subscale; HAM-D<sub>17</sub>: Hamilton Depression Scale, 17 items; BDI: Beck's Depression Inventory; SD: Standard deviation; ECT: Electroconvulsive therapy

**Table 2.** Risk of bias in trials assessing exercise for patients diagnosed with depression

Author, Year of publication	Sequence generation	Allocation concealment	Blinding of participants and trial personnel assessors	Blinding of outcome assessors	Incomplete outcome data	Selective outcome reporting	For profit bias	Other bias	Comment on 'Other bias'
Klein 1985	Unclear	Unclear	High	High	High	Low	Low	Low	
Martinsen 1985	Unclear	Unclear	High	High	High	Low	High	Low	
Epstein 1986	Unclear	Unclear	High	High	High	Low	Unclear	High	Baseline difference
Doyne 1987	Unclear	Unclear	High	Low	High	Low	Unclear	High	Baseline difference
Veale 1992	Unclear	Unclear	High	High	High	Low	Low	High	Baseline difference
Singh 1997	Low	Unclear	High	Low	Low	Low	Low	High	Baseline difference
Blumenthal 1999	Unclear	Unclear	High	Low	High	Low	High	Low	
Mather 2002	Low	Low	High	Low	High	Low	Low	Low	
Dunn 2005	Low	Low	High	Low	High	High	High	Low	
Singh 2005	Low	Low	High	Low	High	Low	Unclear	Low	
Pilu 2007	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Viera 2007	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Blumenthal 2007	Low	Low	High	Low	High	High	Low	Low	
Krogh 2009	Low	Low	High	Low	Low <sup>1</sup>	High	High	High	Baseline difference
Mota-Pereira 2011	Unclear	Unclear	High	Low	High	Low	High	High	Baseline difference
Krogh 2012	Low	Low	High	Low	Low	Low	Low	Low	
Chalder 2012	Low	Low	High	High	Low	Low	Low	Low	
Fang 2013	Unclear	Unclear	High	Unclear	Unclear	High	Unclear	Low	
Huipeng 2013	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Cassandra 2014	Low	Unclear	High	Low	High	Low	Low	Low	
Danielsson 2014	Unclear	Low	High	Low	High	Low	Low	Low	
Pfaff 2014	Low	Low	High	Low	Low <sup>1</sup>	Low	Low	High	Baseline difference
Guifeng 2015	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Jinchun 2015	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Schuch 2015	Unclear	Low	High	Low	Low	Low	Low	Low	
Kerling 2015	Unclear	Unclear	High	Unclear	Low	Low	Low	Low	
Belvederi 2015	Low	Low	High	Low	High	Low	Low	High	Post-hoc sample size
Carneiro 2015	Unclear	Low	High	High	Unclear	Low	Low	Low	
Doose 2015	Unclear	Unclear	High	High	High	Low	Low	High	No sample size calc.
Salehi 2016	High	High	High	Low	Unclear	Low	Low	High	Baseline difference
Legrand 2016	Low	High	High	High	High	Low	Unclear	Low	

**Table 3.** Heterogeneity of effect estimates for trials assessing the effect of exercise for patients diagnosed with depression explored by comparing sub-groups

Subgroups	Number of Trials (participants)	Random effects meta-analysis SMD (95% CI., p, I <sup>2</sup> )	Subgroup explains heterogeneity P value
<b>Risk of bias</b>			
Less than high risk of bias <sup>1</sup>	4 (530)	-0.11 (-0.41 to 0.18; P=0.45; I <sup>2</sup> = 62%)	<0.001
High risk of bias	27 (1801)	-0.85 (-1.10 to -0.60; P<0.001; I <sup>2</sup> = 82%)	
<b>Age</b>			
Old (>59 years)	5 (492)	-0.77 (-1.34 to -0.19; P=0.009; I <sup>2</sup> = 87%)	0.99
Young (<59 years)	26 (1839)	-0.76 (-1.01 to -0.51; P<0.001; I <sup>2</sup> = 83%)	
<b>Exercise context</b>			
Group exercise	24 (1729)	-0.79 (-1.06 to -0.52; P<0.001; I <sup>2</sup> = 85%)	0.72
Individual exercise	7 (602)	-0.68 (-1.17 to -0.20; P=0.005; I <sup>2</sup> = 79%)	
<b>Duration</b>			
Less than 10 weeks	14 (691)	-0.93 (-1.11 to -0.88; P<0.001; I <sup>2</sup> = 19%)	0.05
10 weeks or more	17 (1640)	-0.58 (-0.88 to -0.28; P<0.001; I <sup>2</sup> = 86%)	
<b>Attention control</b>			
Attention control	7 (609)	-0.71 (-1.27 to -0.16; P=0.01; I <sup>2</sup> = 89%)	0.99
Waitlist	2 (47)	-0.67 (-2.48 to 1.13; P=0.47; I <sup>2</sup> = 88%)	
<b>Pharmacotherapy</b>			
Add-on	11 (734)	-0.92 (-1.38 to -0.46; P<0.001; I <sup>2</sup> = 86%)	0.82
No medication	6 (318)	-0.82 (-1.58 to -0.06; P=0.03; I <sup>2</sup> = 88%)	
<b>Somatic comorbidity</b>			
Somatic co-morbidity	0	N/A	
No co-morbidity	31 (2331)	N/A	
<b>Minor depression</b>			
Incl. minor depression	6 (350)	-0.90 (-1.65 to -0.15; P=0.02; I <sup>2</sup> = 86%)	0.67
No minor depression	25 (1981)	-0.73 (-0.97 to -0.49; P<0.001; I <sup>2</sup> = 88%)	
<b>Patient setting</b>			
Inpatients	10 (549)	-0.88 (-1.07 to -0.70; P<0.001; I <sup>2</sup> = 6%)	0.26
Outpatients	21 (1782)	-0.69 (-0.98 to -0.41; P<0.001; I <sup>2</sup> = 85%)	
<b>Trial size</b>			
Trials n ≤ 52	15 (479)	-1.30 (-1.74 to -0.86; P<0.001; I <sup>2</sup> = 77%)	<0.001
Trials n > 52	16 (1852)	-0.40 (-0.60 to -0.19; P<0.001; I <sup>2</sup> = 76%)	

**Table 4.** Summary of findings

## Exercise compared to control or treatment as usual for depression

Patient or population: depression

Setting: In- or out-patients

Intervention: exercise

Comparison: control or treatment as usual

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control or treatment as usual	Risk with exercise				
Severity of depression	-	<b>0.74 SMD lower</b> (0.51 lower to 0.96 lower)	-	2419 (31 RCTs)	⊕○○○ VERY LOW <sup>1</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.
Lack of remission	<b>Study population</b>		<b>RR 0.78</b> (0.68 to 0.90)	1639 (19 RCTs)	⊕○○○ VERY LOW <sup>2</sup>	Remission is, with minor variations, defined as not full-filling the criteria for depression.
	646 per 1000	<b>504 per 1000</b> (426 to 594)				
Serious adverse events	<b>Study population</b>		<b>RR 2.21</b> (0.24 to 20.21)	335 (3 RCTs)	⊕⊕○○ LOW <sup>3</sup>	
	0 per 1000	<b>0 per 1000</b> (0 to 0)				
Quality of life	-	<b>0.43 SMD higher</b> (0.04 lower to 0.91 higher)	-	901 (8 RCTs)	⊕○○○ VERY LOW <sup>4</sup>	Quality of life was assessed using a number of different methods. Higher score indicates improved quality of life. Seven of 24 trials reported on this outcome
Depression severity after the intervention	<b>Study population</b>		<b>RR 0.95</b> (0.82 to 1.11)	777 (5 RCTs)	⊕⊕○○ LOW <sup>6</sup>	
	469 per 1000	<b>446 per 1000</b> (385 to 521)				
Depression severity. Restricted to trials with less than high risk of bias.	-	<b>0.11 SMD lower</b> (0.41 lower to 0.18 higher)	-	530 (4 RCTs)	⊕⊕○○ LOW <sup>7</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.

\*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). CI: Confidence interval; SMD: Standardised mean difference; RR: Risk ratio

### GRADE Working Group grades on evidence

**quality:** We are very confident that the true effect lies close to that of the estimate of the effect

**quality:** We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

**quality:** Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of the effect

High

Moderate

Low

Very low:

1. Downgraded by 3: risk of bias, inconsistency and publication bias
2. Downgraded by 3: risk of bias, inconsistency and publication bias
3. Downgraded by 2: imprecision and publication bias
4. Downgraded by 3: risk of bias, inconsistency and imprecision
5. Downgraded by 2: risk of bias and imprecision
6. Downgraded by 2: risk of bias and imprecision
7. Downgraded by 2: inconsistency and imprecision

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	1
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ for each meta-analysis).	9



# PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	11
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig. 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Fig 3-fig8
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Table 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Table 3
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	21
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	22
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	26

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org).

Page 2 of 2

BMJ Open: first published as 10.1136/bmjopen-2016-014820 on 18 September 2017. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

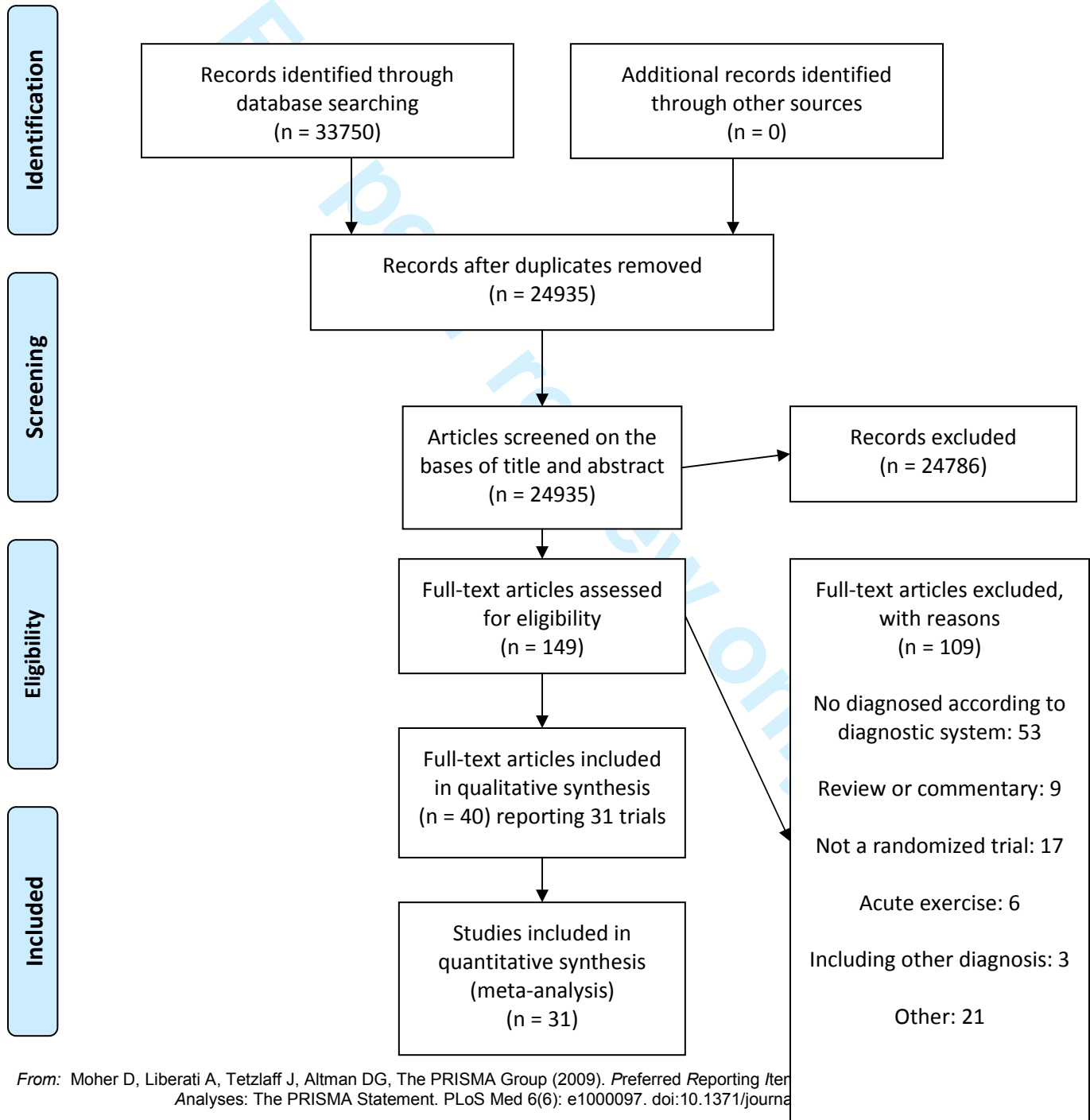


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplemental**

Figure S1. Flow diagram for identification of trials assessing the effects of exercise for patients with depression.



For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).

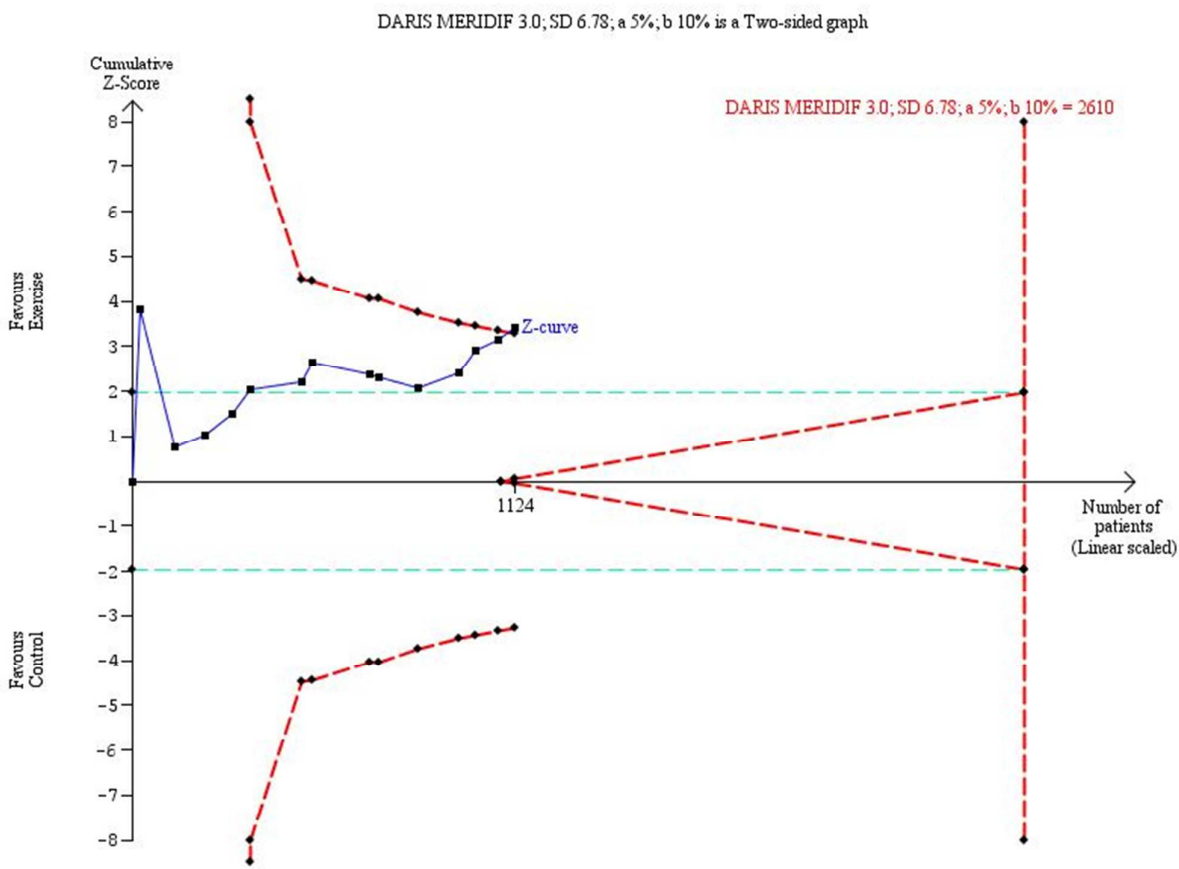
For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S2. Trial Sequential Analysis and required information size for the effect of exercise for depressive symptoms including twelve trials reporting on HAM-D<sub>17</sub>.

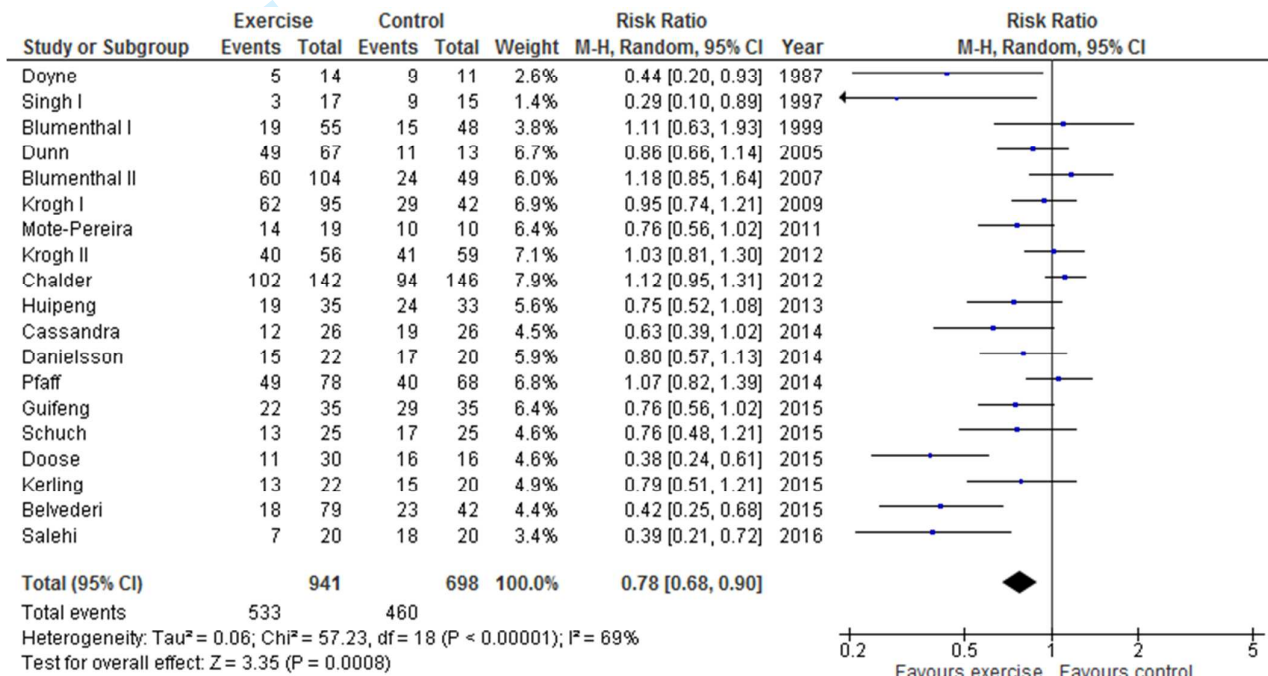


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S3. Effect of exercise on lack of remission for patients diagnosed with depression

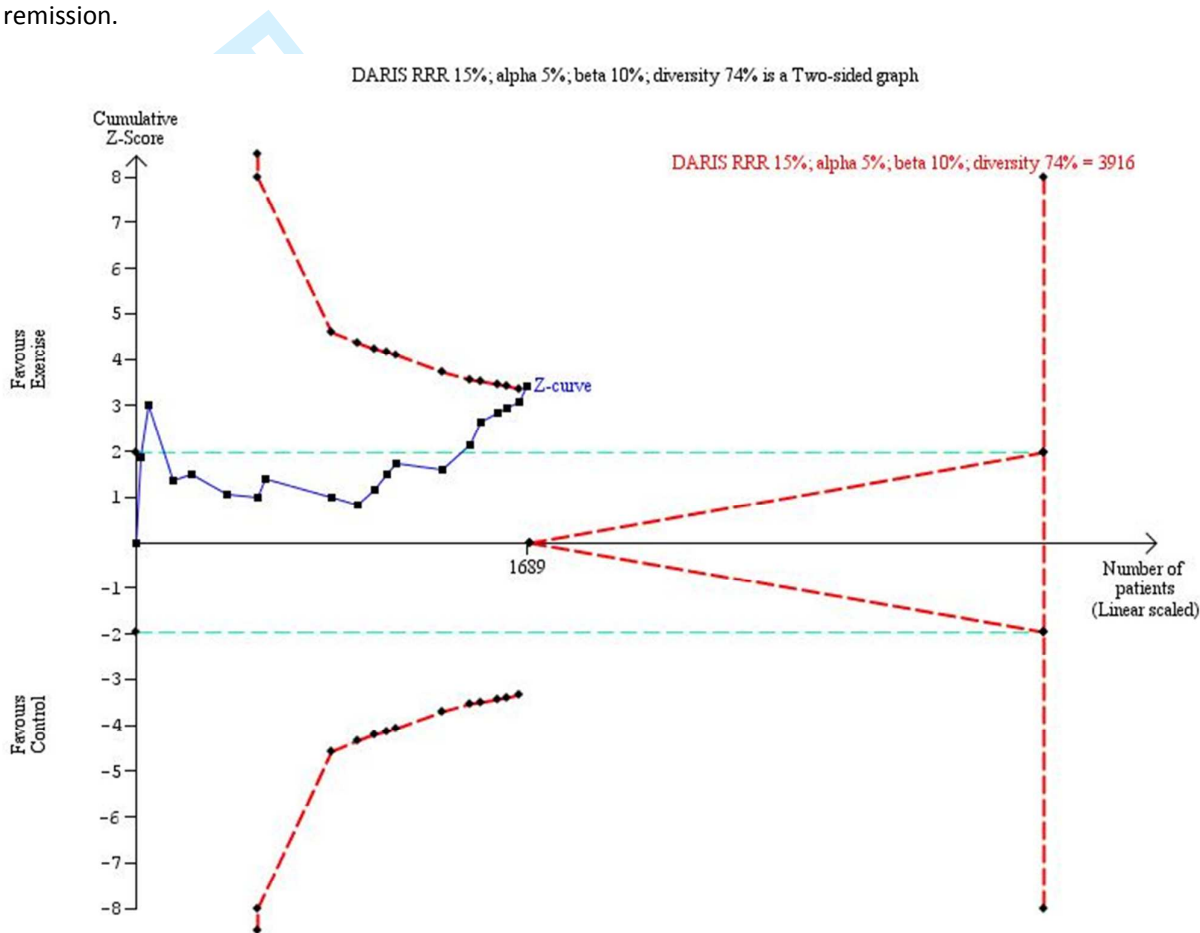


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S3. Trial Sequential Analysis and required information size for the effect of exercise on lack of remission.

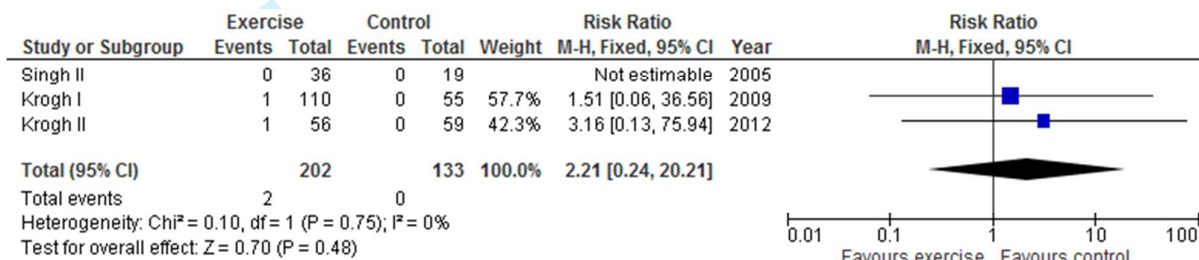


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplementary Figure S5

**Figure S5.** Effect of exercise on risk of serious adverse events for patients diagnosed with depression

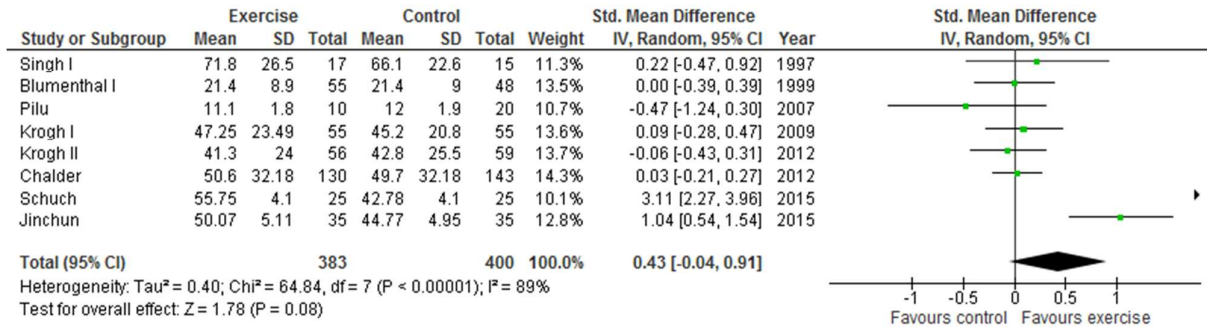


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

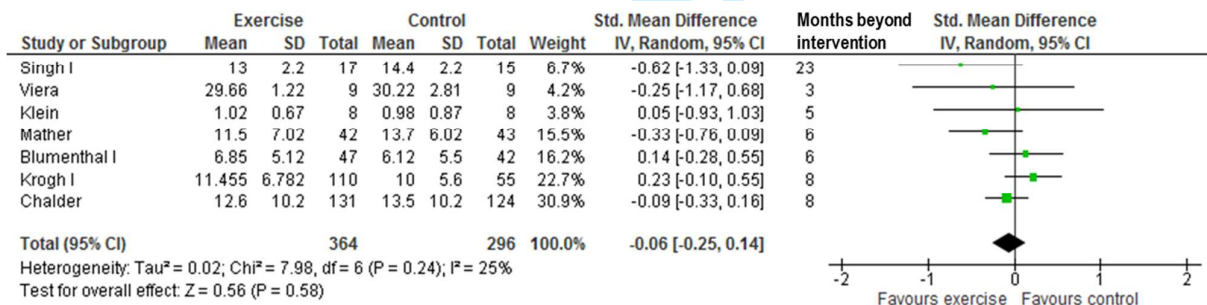
Supplementary Figure S6-S8

Figure S6. The effect of exercise on quality of life in patients diagnosed with depression

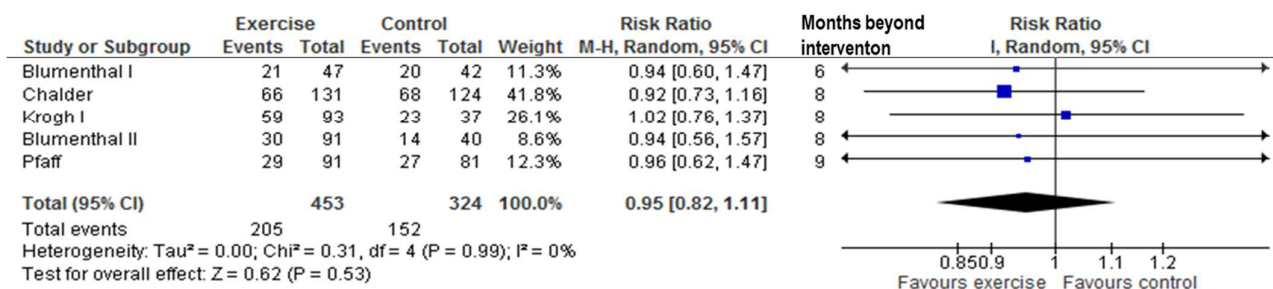


Quality of life was assessed using different scales: Singh I and Chalder used the SF-36, Blumenthal used Life Satisfaction Index, Pilu and Schuch used the WHOQOL, Krogh I and Krogh II used the WHO-Five Well-being Scale, and Jinchun used the GQOLI-74.

Figure S7. The effect of exercise on depression severity after the intervention in patients diagnosed with depression



**Figure S8.** The effect of exercise on risk of lack of remission after the intervention in patients diagnosed with depression



Or peer review only

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplementary Table S1

**Table S1.** Imputation of missing data for trials assessing exercise for patients diagnosed with depression

Outcome	Result from review	Best/worse (1SD)	Best/worse (2SD)	Worse/best (1SD)	Worse/best (2SD)
Depression SMD (95% CI)	-0.74 (-0.96 to -0.51) p < 0.001; I <sup>2</sup> = 83%	-0.85 (-1.10 to -0.60) p < 0.001; I <sup>2</sup> = 87.2%	-0.85 (-1.11 to -0.60) p < 0.001; I <sup>2</sup> = 87.9%	-0.66 (-0.90 to -0.40) p < 0.001; I <sup>2</sup> = 85.4%	-0.61 (-0.84 to -0.38) p < 0.001; I <sup>2</sup> = 85.5%
		<b>Good Outcome</b>	<b>Poor outcome</b>	<b>Good/poor outcome</b>	<b>Poor/good outcome</b>
Lack of remission (95% CL)	RR 0.78 (0.68 to 0.90) p < 0.001; I <sup>2</sup> = 69%	RR 0.75 (0.64 to 0.89) p = 0.0008; I <sup>2</sup> = 73%	RR 0.88 (0.83 to 0.94) p = 0.0002; I <sup>2</sup> = 69%	RR 0.71 (0.61 to 0.81) p < 0.001; I <sup>2</sup> = 68%	RR 0.86 (0.71 to 1.04) p = 0.12; I <sup>2</sup> = 83%
Serious adverse events (95% CL)	RR 2.21 (0.24 to 20.21) p = 0.48; I <sup>2</sup> = 0%	RR 2.19 (0.23 to 20.76) p = 0.50; I <sup>2</sup> = 50%	RR 0.92 (0.37 to 2.30) p = 0.86; I <sup>2</sup> = 60%	RR 0.08 (0.02 to 0.34) p = 0.001; I <sup>2</sup> = 5.4%	RR 19.17 (2.64 to 139.2) p = 0.004; I <sup>2</sup> = 0%

Imputation of missing data for continuous outcome: 'best-worst' - assumed that all participants lost to follow-up in the intervention group had a beneficial outcome (the group mean minus 1 or 2 SD), and all participants lost to follow-up in the placebo group have had a harmful outcome (the group mean plus 1 SD and 2 SD). The reverse 'worst-best-case' scenario is the reverse of the 'best-worst' scenario.

Missing data for the 'remission' outcome was imputed according to the following scenarios: 1) poor outcome analysis: none of the drop-outs/participants lost from both arms experienced the outcome; 2) good outcome analysis: all of the drop-outs/participants lost from both arms experienced the outcome; 3) extreme case analysis favouring the experimental intervention, all of the drop-outs/participants lost from the experimental arm, but none of the drop-outs/participants lost from the control arm experienced the outcome; and 4) extreme case analysis favouring the control: all drop-outs/participants lost from the experimental arm, but none from the control arm experienced the outcome. Missing data for 'serious adverse events' was calculated with the reverse assumptions.



Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplementary Table S2

**Table S2.** Trials characteristics for exploration of heterogeneity in trials assessing the effect of exercise in patients diagnosed with depression

Trial	Lower risk of bias	Age > 60	Group vs. individual	Duration	Attention control waitlist	Exercise as add on to drugs vs. exercise alone	Within-study dose exercise	Somatic disease vs. only MD	Trial Includes minor depression
Klein 1985	No	Young	Individual	12 weeks	Other	Exercise alone	No	No	Yes
Martinsen 1985	No	Young	Group	9 weeks	Attention control	Unclear	No	No	No
Epstein 1986	No	Young	Group	8 weeks	Waitlist	Unclear	No	No	Yes
Doyne 1987	No	Young	Individual	8 weeks	Waitlist	Exercise alone	No	No	Yes
Veale 1992	No	Young	Group	12 weeks	Other	Unclear	No	No	No
Singh 1997	No	Old	Group	10 weeks	Attention control	Exercise alone	No	No	Yes
Blumenthal 1999	No	Young	Group	16 weeks	Other	Add on	No	No	No
Mather 2002	No	Old	Group	10 weeks	Attention control	Add on	No	No	No
Dunn 2005	No	Young	Individual	12 weeks	Attention control	Exercise alone	Yes	No	No
Singh 2005	No	Old	Group	8 weeks	Other	Exercise alone	Yes	No	Yes
Pilu 2007	No	Young	Group	24 weeks	Other	Add on	No	No	No
Viera 2007	No	Young	Group	12 weeks	Other	Add on	No	No	No
Blumenthal 2007	No	Young	Group	16 weeks	Other	Add on	No	No	No
Krogh 2009	Yes	Young	Group	16 weeks	Attention control	No	No	No	No
Mota-Pereira 2011	No	Young	Group	12 weeks	Other	Add on	No	No	No
Krogh 2012	Yes	Young	Group	12 weeks	Attention control	Exercise alone	No	No	No
Chalder 2012	No	Young	Individual	32 weeks	Other	No	No	No	No
Fang 2013	No	Young	Group	6 weeks	Attention control	No	Yes	No	No
Huipeng 2013	No	Young	Group	6 weeks	Other	No	No	No	No
Cassandra 2014	No	Young	Group	3 weeks	Other	Add on	No	No	No
Danielsson 2014	No	Young	Group	10 weeks	Other	Add on	No	No	No
Pfaff 2014	Yes	Old	Group	12 weeks	Other	No	No	No	Yes
Guifeng 2015	No	Young	Group	8 weeks	Other	No	No	No	No
Jinchun 2015	No	Young	Group	8 weeks	Other	No	No	No	No
Schuch 2015	Yes	Young	Individual	2 weeks	Other	No	No	No	No
Kerling 2015	No	Young	Group	6 weeks	Other	No	No	No	No

1	Belvederi	No	Old	Group	24 weeks	Other	Add on	Yes	No	No
2	2015									
3	Carneiro	No	Young	Group	16 weeks	Other	Add on	No	No	No
4	2015									
5	Doose	No	Young	Group	8 weeks	Other	No	No	No	No
6	2015									
7	Legrand	No	Young	Individual	10 days	Other	No	No	No	No
8	2016									
9	Salehi	No	Young	Individual	4 weeks	Other	Add on	No	No	No
10	2016									

For peer review only

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplementary Table

**Table S3.** Heterogeneity of effect estimates for trials assessing the effect of exercise for patients diagnosed with depression on lack of remission.

Subgroups	Number of Trials (participants)	Random effects meta-analysis RR (95% CI., p, I <sup>2</sup> )	Subgroup explains heterogeneity P value
<b>Risk of bias</b>			
Less than high risk of bias <sup>1,2</sup>	2 (165)	0.95 (0.74 to 1.23; p = 0.70; I <sup>2</sup> = 20%)	0.18
High risk of bias	17 (1474)	0.77 (0.64 to 0.92; p = 0.003; I <sup>2</sup> = 75%)	
<b>Age</b>			
Old (>59 years)	3 (299)	0.61 (0.21 to 1.02; p = 0.37; I <sup>2</sup> = 91%)	0.62
Young (<59 years)	16 (1340)	0.81 (0.70 to 0.93; p = 0.003; I <sup>2</sup> = 64%)	
<b>Exercise context</b>			
Group exercise	14 (1156)	0.80 (0.66 to 0.96; p = 0.02; I <sup>2</sup> = 72%)	0.69
Individual exercise	5 (483)	0.74 (0.52 to 1.04; p = 0.08; I <sup>2</sup> = 77%)	
<b>Duration</b>			
Less than 10 weeks	8 (393)	0.63 (0.51 to 0.77; p < 0.001; I <sup>2</sup> = 40%)	0.004
10 weeks or more	11 (1246)	0.93 (0.78 to 1.10; p = 0.39; I <sup>2</sup> = 69%)	
<b>Attention control</b>			
Attention control	4 (364)	0.91 (0.73 to 1.12; p = 0.38; I <sup>2</sup> = 42%)	0.07
Waitlist	1 (25)	0.44 (0.21 to 0.93; p = 0.03; I <sup>2</sup> = 0%)	
<b>Pharmacotherapy</b>			
Add-on	7 (540)	0.72 (0.54 to 0.96; p = 0.03; I <sup>2</sup> = 69%)	0.62
No medication	4 (252)	0.75 (0.52 to 1.09; p = 0.13; I <sup>2</sup> = 66%)	
<b>Somatic comorbidity</b>			
Somatic co-morbidity	0	N/A	
No co-morbidity	19 (1639)	N/A	
<b>Minor depression</b>			
Incl. minor depression	3 (203)	0.63 (0.21 to 1.89; p = 0.41; I <sup>2</sup> = 87%)	0.69
No minor depression	16 (1436)	0.79 (0.68 to 0.92; p = 0.002; I <sup>2</sup> = 69%)	
<b>Patient setting</b>			
Inpatients	6 (322)	0.71 (0.60 to 0.84; p < 0.001; I <sup>2</sup> = 0%)	0.21
Outpatients	13 (1317)	0.84 (0.69 to 1.01; p = 0.07; I <sup>2</sup> = 77%)	
<b>Trial size</b>			
Trials n ≤ 52	9 (358)	0.62 (0.50 to 0.76; p < 0.001; I <sup>2</sup> = 45%)	0.002
Trials n > 52	10 (1281)	0.95 (0.80 to 1.12; p = 0.52; I <sup>2</sup> = 68%)	

<sup>1</sup>Trials potentially having less bias than trials with high risk of bias.

# BMJ Open

## Exercise for patients with major depression: a systematic review with meta-analysis and Trial Sequential Analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-014820.R1
Article Type:	Research
Date Submitted by the Author:	03-Apr-2017
Complete List of Authors:	Krogh, Jesper; Kobenhavns Universitet Sundhedsvidenskabelige Fakultet, Mental Health Centre Copenhagen Speyer, Helene; University of Copenhagen Hjorthøj, Carsten; Copenhagen University Hospital, Mental Health Center Copenhagen Gluud, Christian; Copenhagen University Hospital Rigshospitalet, The Copenhagen Trial Unit, Centre for Clinical Intervention Research Nordentoft, Merete; University of Copenhagen, Mental Health Centre Copenhagen
<b>Primary Subject Heading</b>:	Mental health
Secondary Subject Heading:	Sports and exercise medicine, Evidence based practice
Keywords:	Exercise, MENTAL HEALTH, Systematic review, Meta-analysis, Evidence based medicine, Randomised clinical trials

SCHOLARONE™  
Manuscripts

Only

1  
2  
3  
4 **Exercise for patients with major depression: a systematic review with meta-**  
5 **analysis and Trial Sequential Analysis**  
6

7  
8 Jesper Krogh<sup>1</sup>, Carsten Hjorthøj<sup>1</sup>, Helene Speyer<sup>1</sup>, Christian Gluud<sup>1</sup>, Merete  
9 Nordentoft<sup>1</sup>  
10

11  
12  
13 <sup>1</sup> Mental Health Centre Copenhagen. Faculty of Health Sciences. University of Copenhagen, Copenhagen,  
14 Denmark.

15  
16 <sup>2</sup> Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University  
17 Hospital, Copenhagen, Denmark.  
18  
19

20  
21  
22  
23 **Corresponding author**

24 Jesper Krogh, [Jesper.krogh@dadlnet.dk](mailto:Jesper.krogh@dadlnet.dk)

25  
26 Mental Health Centre Copenhagen, Kildegårdsvej 28, Building 15. DK-2900, Hellerup, Denmark  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

### Objectives

To assess the benefits and harms of exercise in patients with depression.

### Design

Systematic review

### Data sources

Bibliographical databases were searched until the 17<sup>th</sup> of April, 2016.

### Eligibility criteria and outcomes

Eligible trials were randomised clinical trials assessing the effect of exercise in participants diagnosed with depression. Primary outcomes were depression severity, lack of remission, and serious adverse events (e.g. suicide) assessed at the end of the intervention. Secondary outcomes were quality of life and adverse events such as injuries, as well as assessment of depression severity and lack of remission during follow-up after the intervention.

### Results

Thirty-one trials enrolling 2419 participants were included. The effect of exercise versus control on depression severity was -0.74 standardised mean difference (SMD) (95% CI -0.96 to -0.51;  $P < 0.001$ ; GRADE: very low quality). Restricting this analysis to the four trials that seemed less affected of bias, the effect vanished to -0.11 SMD (-0.41 to 0.18;  $P = 0.45$ ; GRADE: low quality). Exercise decreased the relative risk of no remission to 0.78 (0.68 to 0.90;  $P < 0.001$ ; GRADE: very low quality). Restricting this analysis to the two trials that seemed less affected of bias, the effect vanished to 0.95 (0.74 to 1.23;  $P = 0.78$ ). Trial Sequential Analysis excluded random error when all trials were analysed. Sub-group analyses found that

1  
2  
3  
4 trial size and intervention duration were inversely associated with effect size for both depression severity  
5  
6 and lack of remission. There was no significant effect of exercise on secondary outcomes.  
7  
8

### 9 **Conclusions**

10  
11 Trials with less risk of bias suggested no antidepressant effects of exercise and there were no significant  
12  
13 effects of exercise on quality of life, depression severity, or lack of remission during follow-up. Data for  
14  
15 serious adverse events and adverse events was scarce not allowing conclusions for these outcomes.  
16  
17

### 18 **Systematic review registration**

19  
20  
21 The protocol was published in the journal Systematic Reviews: 2015; 4:40  
22  
23

24 **DOI:** 10.1186/s13643-015-0030-6.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Article Summary

### Strengths and limitations of this study

- The protocol for this review has previously been published
- Using meta-regression analysis, trial sequential analysis and the GRADE system the conclusions from this review is based on a firm and transparent platform
- Based on an extensive literature search, this review included 31 trials allocating more than 2000 participants to exercise or control interventions
- The effect estimates are largely based on trials at high risk of bias
- Effect estimates from included trials had considerable heterogeneity



## Introduction

Depression is a common disorder affecting up to 17% of the population during their lifetime.<sup>1,2</sup> Based on data from the World Health Organisation, depression is ranked as the second largest health-care problem globally, in terms of years lived with disability.<sup>3</sup> Depending on its severity, depression is often treated using psychotherapy, antidepressants, or a combination of both. However, the clinical benefits of antidepressants<sup>4,5</sup> and psychotherapy<sup>6-8</sup> has been challenged. Both treatments are costly in terms of time and money and may also have adverse effects. Compliance with antidepressant treatment is poor; the dropout rate in clinical trials is reported to be between 12% and 40% within the initial 6 to 8 weeks of treatment.<sup>4,9</sup>

The weakness of evidence for the beneficial effect of current interventions, along with problems related to low compliance and harms, has resulted in an interest in using alternative interventions. The use of exercise as an intervention has attracted considerable attention, and various forms of exercise varying in intensity have been assessed in a number of randomised clinical trials to test their effectiveness as a treatment for patients with depression. In 2011, we published a meta-analysis of randomised clinical trials examining the effect of exercise on depressive symptoms in patients with clinical depression.<sup>10</sup> The results suggested that referring patients with clinical depression to exercise programs was associated with a small to moderate effect on depressive symptoms. However, restricting the analysis to three trials with a low risk of bias, the effect estimate was non-significant. Since 2011, other reviews have been published on the effect of exercise on depressive symptoms,<sup>11</sup> in older people,<sup>12</sup> and in patients with chronic illnesses.<sup>13</sup> However, none of these reviews addressed the specific population of adults diagnosed with major depression according to valid diagnostic criteria, such as the International Classification of Diseases<sup>14</sup> or the Diagnostic and Statistical Manual of Mental Disorders.<sup>15</sup> The reviews contained a number of trials that included volunteers who were defined as being depressed on the basis of psychometric testing (for example, Beck

1  
2  
3  
4 Depression Inventory<sup>16</sup>), as opposed to individuals with a clinical diagnosis of major depression.

5  
6 Furthermore, several randomised clinical trials investigating the effect of exercise in clinically depressed  
7  
8 individuals have been published since our 2011 review.<sup>10</sup>  
9

10  
11  
12  
13  
14 The objectives of the present systematic review are to investigate the beneficial and harmful effects of  
15  
16 exercise, in terms of severity of depression, lack of remission, quality of life, and suicide versus controls  
17  
18 with or without co-interventions in adults with a clinical diagnosis of major depression. The current  
19  
20 systematic review differs from our previous review in a number of aspects.<sup>10</sup> We only considered trials  
21  
22 including participants diagnosed with depression according to a validated diagnostic system. We also  
23  
24 included trials including participants with somatic co-morbidity, e.g., cancer or diabetes. The harmful  
25  
26 effects of exercise interventions are also addressed, the intervention effects being assessed according to  
27  
28 the grading of recommendations assessment, development, and evaluation (GRADE) framework, and  
29  
30 bibliographical searches have been extended to include a Chinese and a South-American database until  
31  
32 2016.  
33  
34  
35  
36  
37  
38

### 39 **Methods/design**

40  
41  
42 The protocol for this review has previously been published.<sup>17</sup>  
43  
44  
45  
46

### 47 **Search strategy**

48  
49  
50 The following bibliographical databases was searched from April 2015 until the 17<sup>th</sup> of April, 2016:  
51  
52 CENTRAL, MEDLINE, EMBASE, Science Citation Index (Web of Science), LILACS, and Wanfang using medical  
53  
54 subject headings (MeSH or similar) when possible or text word terms: depression, depressive disorder and  
55  
56 exercise, aerobic, non-aerobic, physical activity, physical fitness, walking, jogging, running, bicycling,  
57  
58  
59  
60

1  
2  
3  
4 swimming, strength, or resistance. Please see supplementary material (S1) for an example of a  
5  
6 bibliographical search.  
7  
8  
9

### 10 11 **Trial selection**

12  
13  
14 One investigator (JK) examined titles and abstracts to remove obviously irrelevant reports. Two  
15  
16 investigators (JK + HS) examined full text reports and abstracts determining compliance with inclusion  
17  
18 criteria. A trial was considered eligible if it was a randomised clinical trials including participants diagnosed  
19  
20 as having major depression according to a valid and recognised diagnostic system (that is, Research  
21  
22 Diagnostic Criteria (RDC),<sup>18</sup> International Classification of Diseases (ICD),<sup>14</sup> or Diagnostic and Statistical  
23  
24 Manual of Mental disorders (DSM)<sup>15</sup>) and included participants aged >17 years. Abstracts and full text  
25  
26 reports were included.  
27  
28  
29

30  
31  
32  
33 Trials were excluded if they measured depression immediately after a single bout of exercise, compared  
34  
35 one form of exercise versus another, or compared different exercise intensities without including a control  
36  
37 group. The trials had to allocate participants to an exercise intervention versus a control group (that is,  
38  
39 exercise versus a control group receiving no intervention or treatment as usual or an attention control  
40  
41 using light exercise) or using exercise as an add-on-treatment (that is, exercise plus usual treatment in the  
42  
43 experimental group versus usual treatment alone in the control group). Exercise intervention was defined  
44  
45 as a systematic physical intervention with the intention to increase muscle strength and/or cardiovascular  
46  
47 fitness, e.g., running, swimming or weight lifting. In case of attention control, it should specifically be  
48  
49 mentioned by the authors of the trial report that the intervention was intended as a control intervention.  
50  
51  
52

### 53 54 55 **Outcomes**

1  
2  
3  
4 The primary outcomes were 1) depressive symptoms measured on a continuous scale assessed at the end  
5  
6 of the intervention; 2) lack of remission, that is, a binary outcome of the proportion of participants in each  
7  
8 intervention group of the trial who did not obtain remission at the end of the intervention according to the  
9  
10 authors' own definition; and 3) serious adverse events defined according to ICH-GCP as any untoward  
11  
12 medical occurrence that was life threatening, resulted in death or persistent or significant disability (ICH-  
13  
14 GCP 1997).<sup>19</sup> Serious adverse events accordingly include suicide attempts as well as suicides. The secondary  
15  
16 outcomes were quality of life, non-serious adverse events (e.g., muscle injuries) as well as depressive  
17  
18 symptoms and lack of remission assessed after the intervention.  
19  
20  
21  
22  
23  
24

### 25 **Data extraction**

26  
27 Two authors (JK, HS) independently extracted data using a pre-piloted structured form. Any discrepancies  
28  
29 in the data extraction or inclusion/exclusion of trials was resolved by referring to the original papers. CG or  
30  
31 MN assisted as adjudicator in cases of disagreements. Data extraction included, in addition to outcomes,  
32  
33 information regarding country of origin, number of randomised participants, number of participants  
34  
35 included in efficacy analysis, mean age of participants, diagnostic system, baseline assessment of  
36  
37 depression severity, type of intervention, frequency of intervention, and duration of intervention.  
38  
39 Continuous outcomes were preferred in the following order: post-intervention scores with corresponding  
40  
41 standard deviations (SD), mean change from baseline with SD, mean difference between groups post-  
42  
43 intervention and reported outcomes were preferred to figure's. JK and CH independently performed the  
44  
45 assessment of bias domains. The authors JK, CG, and MN have previously published trial reports assessing  
46  
47 the effect of exercise in participants with depression,<sup>20;21</sup> and to reduce the risk of academic bias two  
48  
49 additional authors were included in the current systematic review (CH, HS).  
50  
51  
52  
53  
54  
55  
56

### 57 **Risk of bias assessment**

1  
2  
3  
4 Definitions in the assessment of bias risk of a trial was conducted according to the Cochrane Handbook for  
5  
6 Systematic Reviews of Interventions<sup>22</sup> of the following domains: allocation sequence generation, allocation  
7  
8 concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome  
9  
10 data, selective outcome reporting, for-profit bias, and other bias. Trials assessed as having 'low risk of bias'  
11  
12 in all of the above specified domains were considered 'trials at low risk of bias'. Trials assessed as having  
13  
14 'uncertain risk of bias' or 'high risk of bias' in one or more of the above specified domains were considered  
15  
16 trials with 'high risk of bias'. In line with our previous systematic review<sup>10</sup> and the latest Cochrane review on  
17  
18 exercise for depression,<sup>23</sup> trials at low risk of bias in the allocation concealment domain, blinded outcome  
19  
20 assessment domain, and the incomplete outcome data domain were characterised as 'trials potentially  
21  
22 having less risk of bias than other trials at high risk of bias'. Trials assessing the effect of behavioural  
23  
24 interventions are rarely able to mask the allocation, and participants and health care providers are  
25  
26 therefore not blinded. Therefore, we will also report the number of trials at low risk of bias in the remaining  
27  
28 domains.  
29  
30  
31  
32  
33  
34  
35

### 36 **Data synthesis and analysis**

37  
38 In order to be able to include all of the trials in our meta-analysis, estimates of standardised mean  
39  
40 difference (SMD) for each individual trial was carried out. SMD is the mean difference in depression score  
41  
42 between the exercise and control groups divided by the pooled standard deviation. The result is a unit free  
43  
44 effect size. By convention, SMD effect sizes of 0.2, 0.5 and 0.8 are considered small, medium and large  
45  
46 intervention effects.<sup>22</sup> For dichotomous variables, we calculated the risk ratio (RR) with a 95% confidence  
47  
48 interval. It was expected that some trials would have several intervention groups. Data from the  
49  
50 experimental groups was pooled and compared with the data from the control group. In case of  
51  
52 discrepancies between the random-effects model analysis and the fixed-effect model analysis, both results  
53  
54 are reported; otherwise, only results from the random-effects analysis is reported. The degree of  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 heterogeneity was quantified using the I-squared statistic,<sup>24</sup> which can be interpreted as the percentage of  
5  
6 variation observed between the trials attributable to between-trial differences, rather than sampling error  
7  
8 (chance). Heterogeneity was explored by analyses of sub-groups (see below).  
9

10  
11  
12  
13  
14 For the primary outcomes, Trial Sequential Analysis was performed.<sup>25,26</sup> In order to calculate the required  
15  
16 information size and the cumulative Z-curve's eventual breach of relevant trial sequential monitoring  
17  
18 boundaries, the required information size for the primary continuous outcome was based on type I error of  
19  
20 5%, a beta of 10%, the standard error of the meta-analysis, and a minimal difference of three points on the  
21  
22 HAM-D<sub>17</sub>.<sup>17</sup> Post-hoc we calculated the required information size including all trials. This was done by  
23  
24 converting effect estimates from trials reporting other outcome scales into the HAM-D<sub>17</sub> scale as described  
25  
26 by Thorlund et al.<sup>27</sup> In order to calculate the required information size and the cumulative Z-curve's  
27  
28 eventual breach of relevant trial sequential monitoring boundaries, the required information size for lack of  
29  
30 remission was based on type I error of 5%, a beta of 10%, the proportion of participants in the control  
31  
32 group with the outcome, and a relative risk reduction of 15% and 30%.  
33  
34  
35  
36  
37  
38

39 Bayes factors were calculated for all primary outcomes.<sup>28</sup> Low P-values suggest that we can reject the null-  
40  
41 hypothesis. But even a low P-value from a meta-analysis can be misleading if there is also a low probability  
42  
43 that data are compatible with the anticipated intervention effect. In other words, the probability that the  
44  
45 actual measured difference in effect of the compared interventions resulted from an a priori anticipated  
46  
47 'true' difference needs to be considered. For this purpose, it is helpful to calculate the Bayes factor, which  
48  
49 is the ratio of the P-value probabilities of the meta-analysis result divided by the probability of the  
50  
51 anticipated effect, or 'true' effect.<sup>28</sup> As suggested by Jakobsen et al.,<sup>28</sup> a Bayes factor lower than 0.1  
52  
53 together with a low P-value suggest, if bias can be ruled out, that the observed result is compatible with the  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 a priori expected effect. If the Bayes factor is higher than 0.1 the result is not compatible with the a priori  
5  
6 expected effect and the effect may be lower.  
7  
8  
9

10  
11 To assess the potential impact of missing data (incomplete outcome data bias) we did sensitivity analysis of  
12  
13 missing data using the following strategy: a 'best-worst' case scenario was assessed, assuming that all  
14  
15 participants lost to follow-up in the intervention group had a beneficial outcome (the group mean minus 1  
16  
17 standard deviation (SD)), and all those with missing outcomes in the control group have had a harmful  
18  
19 outcome (the group mean plus 1 SD and 2 SD). In addition, the reverse 'worst-best-case' scenario analysis  
20  
21 was also performed.<sup>28</sup> Missing data for the 'lack of remission' outcome was imputed in sensitivity analysis  
22  
23 according to the following scenarios:<sup>29</sup> 1) poor outcome analysis: assuming that all of the drop-  
24  
25 outs/participants lost from both the experimental and the control arms experienced the outcome, including  
26  
27 all randomised participants in the denominator; 2) good outcome analysis: assuming that none of the drop-  
28  
29 outs/participants lost from the experimental and the control arms experienced the outcome, including all  
30  
31 randomised participants in the denominator; 3) extreme case analysis favouring the experimental  
32  
33 intervention ('best-worse' case scenario): none of the drop-outs/participants lost from the experimental  
34  
35 arm, but all of the drop-outs/participants lost from the control arm experienced the outcome, including all  
36  
37 randomised participants in the denominator; and 4) extreme case analysis favouring the control ('worst-  
38  
39 best' case scenario): all of the drop-outs/participants lost from the experimental arm, but none from the  
40  
41 control arm experienced the outcome, including all randomised participants in the denominator.  
42  
43  
44  
45  
46  
47  
48  
49

### 50 **Subgroup analyses**

51  
52 In subgroup analyses, the possible effects of variables on intervention effects on outcomes and  
53  
54 heterogeneity were compared. Trials potentially having less risk of bias (i.e., trials with adequate allocation  
55  
56 concealment, blinded outcome assessment, and intention to treat analysis) were compared to trials at high  
57  
58  
59  
60

1  
2  
3  
4 risk of bias. The effect of age was assessed by comparing trials including older participants (mean age >59  
5  
6 years) to trials including younger participants (mean age <60 years). The effect of type of exercise was  
7  
8 assessed by comparing trials using group exercises compared to trials using individual exercise. The effect  
9  
10 of duration of intervention was assessed by comparing trials with short duration of intervention to trials  
11  
12 with long duration of intervention splitting by the median time of duration. The effect of type of control  
13  
14 group was assessed by comparing trials using attention control to trials with waitlist controls and  
15  
16 comparing trials with exercise as add-on to medication to trials not using any medication. In addition, a  
17  
18 within-study comparison of low-dose exercise versus high-dose exercise in trials using different exercise  
19  
20 intensities was performed. The effect of co-morbid somatic disease was assessed by comparing the effect  
21  
22 estimates from trials including participants with depression compared to trials including participants with  
23  
24 depression in addition to a somatic disease. Publication bias was assessed by visual inspection of a funnel  
25  
26 plot and by Egger's test and if publication bias plausible Duval's and Tweedie's trim and fill procedure was  
27  
28 conducted.<sup>30</sup>  
29  
30  
31  
32  
33  
34  
35

36 We assessed and graded the evidence according to the grading of recommendations assessment,  
37  
38 development, and evaluation (GRADE) for high risk of bias, imprecision, indirectness, heterogeneity, and  
39  
40 publication bias.<sup>31</sup> Based on this assessment, the intervention is graded accordingly: 'high quality'- we are  
41  
42 very confident that the true effect lies close to that of the estimate of the effect; 'moderate quality'- we are  
43  
44 moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the  
45  
46 effect, but there is a possibility that it is substantially different; 'low quality'- our confidence in the effect  
47  
48 estimate is limited: the true effect may be substantially different from the estimate of the effect; 'very low  
49  
50 quality'- we have very little confidence in the effect estimate: the true effect is likely to be substantially  
51  
52 different from the estimate of the effect.<sup>32</sup>  
53  
54  
55  
56  
57  
58  
59  
60



### Deviations from our protocol

Post-hoc we included trials using the Chinese Classification of Mental Disorders (CCMD) as well as a few trials including participants classified as having 'minor depression'. The CCMD system closely adhere to the ICD and DSM systems and have been found highly compatible in field studies, so these studies were included.<sup>33</sup> A few trials included some participants classified as having 'minor depression' according to the trials chosen diagnostic system (e.g., DSM), and it is questionable if these participants have major depression. We therefore decided to include these trials but also to conduct a sub-group analysis exclusively including participants with major depression. To further explore heterogeneity, we post-hoc included sub-group analysis comparing intervention effects in inpatients and outpatients as well as an analysis according to trial size. Trials were divided into small or large trials using the median of total n included in the efficacy analysis. The effect of exercise capacity was post-hoc assessed by comparing trials with a high increase in maximal oxygen uptake (VO<sub>2</sub>max) with studies with lower increase in maximal oxygen uptake. Assessment of exercise capacity was based on the increase of VO<sub>2</sub>max in the intervention groups and trials were stratified to either high or low increase in exercise capacity by median. We did not conduct Trial Sequential Analysis based on a relative risk reduction of 30% of lack of remission as this was an implausible effect.

### Participant involvement

Depressed participants were not involved in this study.

### Results

#### Bibliographical search and trial characteristics

1  
2  
3  
4 The main bibliographical search was conducted the 26<sup>th</sup> of August, 2015 and the final updates were  
5  
6 conducted on the 17<sup>th</sup> of April, 2016. As illustrated in Figure S1, we identified 40 publications reporting the  
7  
8 effect of exercise on depressive symptoms in 31 randomised clinical trials.<sup>20;21;34-72</sup> Four-teen trials were  
9  
10 conducted in Europe,<sup>20;21;39;48;51;52;54;60;64-67;73;74</sup> seven in the U.S.A.,<sup>37;38;42;44;59;63;75</sup> six in Asia,<sup>46;68-72</sup> two in  
11  
12 Australia,<sup>53;57</sup> and two in South-America.<sup>55;62</sup> A total of 2,419 participants were randomised and 2,331 were  
13  
14 included in the efficacy analysis of benefit. 10 trials included inpatients<sup>46;48;55;66;68-73</sup> and five trials included  
15  
16 participants with a mean age above 60 years.<sup>51;53;57;59;60</sup> No trials exclusively included participants with  
17  
18 comorbid somatic disease. Four trials reported the continuous outcome as mean change from baseline in  
19  
20 each group with a corresponding SD,<sup>38;52;64;67</sup> and one trial presented data as mean difference between  
21  
22 groups post-intervention.<sup>39</sup> The remaining trials reported post-scores in each group with corresponding SD.  
23  
24 Please see Table 1 for trial characteristics.  
25  
26  
27  
28  
29  
30

### 31 *Bias risk assessment*

32  
33  
34 Sequence generation was adequate in 12/31 (39%), allocation concealment was adequate in 12/31 (39%)  
35  
36 trials, blinding of participants and trial personnel was adequate in 0/31 (0%), blinded outcome assessment  
37  
38 was performed in 16/31 (52%), low risk of bias in the 'incomplete outcome data' domain was found in  
39  
40 12/31 (39%) trials, selective outcome reporting domain was adequate in 27/31 (87%), for profit bias  
41  
42 domain was adequate in 15/31 (48%) and 21/31 (68%) were free of other bias. All trials were at high risk of  
43  
44 bias. Given the nature of the intervention, no trial had blinded participants or trial personnel, however, two  
45  
46 trials had low risk of bias in all other bias domains.<sup>21;53</sup> Five trials (16%) were sponsored by for profit  
47  
48 organisations: three trials were supported by pharmaceutical companies,<sup>52;73;76</sup> one trial by a company  
49  
50 producing fitness machines,<sup>44</sup> and one trial by an insurance company.<sup>20</sup> According to our a priori defined  
51  
52 criteria, 4/31 (13%) trials potentially had less risk of bias than the other trials at high risk of bias.<sup>20;21;53;55</sup>  
53  
54 Please see Table 2 for details on assessment of risks of bias.  
55  
56  
57  
58  
59  
60

## Primary outcomes

### The effect of exercise on depression severity

All included trials provided a continuous outcome on depression severity for the assessment of the exercise intervention encompassing 2,331/2,419 randomised participants (96.4%). The effect of intervention versus control was a standardised mean difference (SMD) of -0.74 (95% CI -0.96 to -0.51;  $P < 0.001$ ) (Figure 1.). This corresponds to an effect on the HAM-D<sub>17</sub> scale of -4.6 (95% CI -6.0 to -3.2) points.

### Missing data

Missing outcome analysis for depression as a continuous outcome did not markedly change the effect estimates. The least favourable outcome for the exercise intervention was the worse/best outcome analysis using +2 SD resulting in an effect estimate of -0.61 SMD (95% CI -0.84 to -0.37;  $P < 0.001$ ) (Table S1).

### Heterogeneity and subgroup analysis

The  $I^2$  was 82% suggesting substantial heterogeneity. Subgroup analysis revealed that the effect estimates for trials potentially having less risk of bias was -0.11 SMD (95% CI -0.41 to 0.18;  $P = 0.45$ ;  $I^2 = 62\%$ ) compared to that of the trials at high risk of bias -0.85 SMD (95% CI -1.10 to -0.60;  $P < 0.001$ ;  $I^2 = 82\%$ ) (test of sub-group difference,  $P = 0.0002$ ). In addition, trials including 52 participants or less had a pooled estimate of -1.30 SMD (95% CI -1.74 to -0.86;  $P < 0.001$ ;  $I^2 = 77\%$ ) compared to that of larger trials of -0.40 SMD (95% CI -0.60 to -0.19;  $P < 0.001$ ;  $I^2 = 76\%$ ) (test of sub-group difference,  $P < 0.001$ ). Trials of short duration of intervention (less than 10 weeks) had a SMD of -0.93 (95% CI -1.11 to -0.88;  $P < 0.001$ ;  $I^2 = 19\%$ ) compared to trials with longer duration of intervention, -0.58 SMD (95% CI -0.88 to -0.28;  $P < 0.001$ ;  $I^2 = 86\%$ ) (test of sub-group difference,  $P = 0.05$ ). Effect estimates from trials including participants with minor

1  
2  
3  
4 depression compared to trials exclusively including participants with major depression did not differ (test of  
5  
6 sub-group difference,  $P = 0.67$ ).  
7  
8

9  
10  
11 Four trials allocated 206 participants to different exercise intensities/doses.<sup>44;57;72;77</sup> Comparing the post-  
12  
13 intervention depression scores for participants allocated to either high intensity/high dose or low  
14  
15 intensity/low dose exercise showed a difference of -0.40 SMD (95% CI -0.67 to -0.12;  $P=0.005$ ;  $I^2 = 0\%$ ) in  
16  
17 favour of high intensity/high dose exercise. As shown in Table 3, no other trial characteristic significantly  
18  
19 explained any of the observed heterogeneity. Please see Table S2 for trial characteristics used to explore  
20  
21 heterogeneity.  
22  
23

#### 24 25 26 27 28 *Trial Sequential Analysis and diversity adjusted required information size*

29  
30  
31 The diversity adjusted required information size for HAM-D<sub>17</sub> as a continuous outcome was calculated  
32  
33 based on our anticipated intervention effect of a minimal relevant difference of 3.0 HDRS points, a standard  
34  
35 deviation of 6.78 points, a risk of type I error of 0.05, a power of 90% and the observed diversity of 92% to  
36  
37 2610 participants. Only 14 trials reported results from HAM-D<sub>17</sub><sup>20;21;37;38;42;43;51;52;54;55;57;67;69;77</sup> with an accrued  
38  
39 1124 participants. As shown in Figure S2, the cumulative Z-curve just crossed the trial sequential  
40  
41 monitoring boundary for benefit. With the aforementioned settings, the pooled estimate is therefore less  
42  
43 likely to be a random finding due to lack of power or multiple testing if bias could be ignored. Post-hoc we  
44  
45 calculated the adjusted required information size for HAM-D<sub>17</sub> including all trials as shown in Figure S3. As  
46  
47 with the original analysis the Z-curve crossed the trial sequential monitoring boundary for benefit  
48  
49 supporting that the pooled estimate is less likely to represent a Type 1 error if bias could be ignored.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Bayes factor

Fourteen trials reported effect estimates using the HAM-D<sub>17</sub>.<sup>20;21;37;38;42;44;51;52;54;62;67;69;77;78</sup> Based on these trials, Bayes factor was calculated ( $\delta = -3.37$ ;  $SE_{\delta} = 0.96$ ;  $\mu_a = -3.0$ ) and was found to be 0.002, which is below the Bayes factor threshold for significance of 0.1, supporting the intervention effect if bias could be ignored.

### Publication bias

Inspection of the funnel-plot (not shown) suggested that small trials with small or no effect of exercise were missing (Figure S4). Egger's test supported the suspicion of publication bias,  $P < 0.00001$ . Using the Duval and Tweedie's trim and fill procedure, the estimate was reduced to -0.28 SMD (95% CI -0.52 to -0.04). This corresponds to an effect on the HAM-D<sub>17</sub> scale of -1.8 (95% CI -3.2 to -0.25).

### The effect of exercise on depression – lack of remission

Nineteen trials, randomising 1825 participants and including 1639 participants (90%) in final analysis reported remission as an outcome.<sup>20;21;37-39;42;44;46;48;52;53;55;59;60;64;67-69;71</sup> Remission post-intervention was defined in various ways: A post-intervention score on the HAM-D<sub>17</sub> less than 8 points,<sup>43;52;55;68;69</sup> not fulfilling the DSM criteria for depression *and* a HAM-D<sub>17</sub> less than 8 points,<sup>20;21;38</sup> not fulfilling the DSM criteria for depression,<sup>37;53;59</sup> a BDI score less than 9 points,<sup>42</sup> a BDI score less than 10 points,<sup>39</sup> a HAM-D<sub>17</sub> score less than 10 points,<sup>77</sup> a MADRS score less than 10 points,<sup>46</sup> a MADRS score less than 10 points *and* a 50% reduction in symptom score,<sup>64</sup> a 75% reduction in HAM-D<sub>24</sub>,<sup>71</sup> a HAM-D<sub>17</sub> score less than 11.28 points *and* a reduction in HAM-D<sub>17</sub> scores  $> 7.74$  points,<sup>67</sup> and one study used MADRS not specifying the cut-off for remission.<sup>48</sup> The RR for lack of remission was 0.78 (95% CI 0.68 to 0.90;  $P=0.0008$ ) in favour of the

1  
2  
3  
4 intervention using a random-effects analysis. The  $I^2$  was 69% suggesting substantial heterogeneity. The  
5  
6 forest plot for the intervention effect on lack of remission is illustrated in Figure S5.  
7  
8  
9

#### 10 11 *Missing data*

12  
13  
14 The scenario in least favour of the intervention was the 'poor' outcome analysis having an effect estimate  
15  
16 of RR 0.88 (95% CI 0.83 to 0.94)  $P=0.0002$ ;  $I^2 = 69\%$ . As shown in Table S1, the remaining scenarios did not  
17  
18 substantially differ from the main analysis.  
19

#### 20 21 *Heterogeneity and subgroup analysis*

22  
23  
24  $I^2$  was 69% for the outcome lack of remission suggesting substantial heterogeneity. For this outcome, only  
25  
26 two trials<sup>21;78</sup> were considered as trials potentially having less risk of bias than the other trials at high risk of  
27  
28 bias. The RR of these two trials was 0.95 (95% CI 0.74 to 1.23;  $P=0.78$ ) compared to 0.77 (96% CI 0.64 to  
29  
30 0.92;  $P=0.003$ ) for trials at high risk of bias, test of subgroup difference,  $P=0.19$ ). Trials including 52  
31  
32 participants or less in their final analysis had a RR 0.62 (95% CI 0.50 to 0.76;  $P<0.001$ ;  $I^2 = 45\%$ ) compared to  
33  
34 0.95 (95% CI 0.80 to 1.12;  $P=0.52$ ;  $I^2 = 68\%$ ) for larger trials (test of sub-group difference,  $P=0.002$ ). Also,  
35  
36 trials with a duration of less than 10 weeks had a RR of 0.63 (95% CI 0.51 to 0.77;  $P<0.001$ ;  $I^2 = 40\%$ )  
37  
38 compared to 0.93 (95% CI 0.78 to 1.10;  $P=0.39$ ;  $I^2 = 69\%$ ) for trials of a longer duration (test of sub-group  
39  
40 difference,  $P=0.004$ ). As shown in Table S3, no other trial characteristic significantly explained any of the  
41  
42 observed heterogeneity. Please see Table S2 for trial characteristics used to explore heterogeneity.  
43  
44  
45  
46  
47  
48  
49

#### 50 51 *Trial Sequential Analysis and diversity adjusted required information size*

52  
53 The diversity adjusted required information size for lack of remission was calculated based on our observed  
54  
55 diversity of 74%, a proportion in the control group with lack of remission of 66%, an anticipated  
56  
57  
58  
59  
60

1  
2  
3  
4 intervention effect of 15% relative risk reduction, a risk of type I error of 0.05% and a power of 90%. As  
5  
6 shown in Figure S6, the cumulative Z curve just crossed the trial sequential monitoring boundary for  
7  
8 benefit. With the aforementioned settings, the pooled estimate is therefore less likely to be a random  
9  
10 finding due to lack of power or multiple testing if bias could be ignored.  
11  
12  
13

#### 14 15 16 17 18 *Bayes factor*

19  
20 Bayes factor was calculate based on the observed relative risk of remission, the associated standard error,  
21  
22 and an anticipated intervention effect of relative increase in number of participants with remission by 15%  
23  
24 ( $\delta = -0.248$ ;  $SE_{\delta} = 0.08$ ;  $\mu_{\delta} = -0.163$ ). Bayes factor was 0.02, which is below the Bayes factor threshold for  
25  
26 significance of 0.1.  
27  
28  
29  
30  
31  
32

#### 33 *Publication bias*

34  
35 Inspection of the funnel-plot (not shown) suggested that small trials with small or no effect of exercise  
36  
37 were missing. Egger's test supported the suspicion of publication bias,  $P=0.002$ . Imputing theoretically  
38  
39 missing studies by the Duval and Tweedie's trim and fill procedure, reduced the estimate of intervention  
40  
41 effect to a relative risk reduction of 0.93 (95% CI 0.79 to 1.11).  
42  
43  
44  
45  
46  
47

#### 48 **The effect of exercise on serious adverse events**

49  
50 Serious adverse events (i.e., death or suicide attempts) were reported in only three trials.<sup>20,21;57</sup> In these  
51  
52 trials, one suicide attempt<sup>21</sup> and one death by suicide<sup>20</sup> were recorded in the intervention groups. The RR  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 for death or suicide in the two trials was 2.21 (95% CI 0.24 to 20.21; P=0.48; I<sup>2</sup> = 0%) as illustrated in Figure  
5  
6 S7.  
7

#### 8 9 10 11 *Missing data*

12  
13  
14 Missing outcome analysis for 'serious adverse events' varied according to missing data scenario: poor  
15  
16 outcome analysis relative risk, 0.92 (95% CI 0.37 to 2.30; P=0.86; I<sup>2</sup> = 60.0%), good outcome analysis, 2.19  
17  
18 (95% CI 0.23 to 20.76; P=0.50; I<sup>2</sup> = 0.0%), best/worst outcome analysis – 0.08 (95% CI 0.02 to 0.34; P=0.001;  
19  
20 I<sup>2</sup> = 5.4%), worst/best outcome analysis 19.17 (95% CI 2.64 to 139.2; P=0.004; I<sup>2</sup> = 0.0%).  
21  
22  
23  
24  
25  
26

#### 27 *Trial Sequential Analysis and Bayes analysis*

28  
29 We decided not to conduct Trial Sequential Analysis or Bayes analysis due to too sparse data.  
30  
31  
32  
33  
34

#### 35 *Publication bias*

36  
37  
38 Only 3/31 trials reported on this outcome and no formal assessment for publication bias was made.  
39  
40 However, the lack of reporting in the vast majority of trials suggest risk publication bias.  
41  
42  
43  
44

#### 45 **Secondary outcomes**

##### 46 47 *The effect of exercise on quality of life*

48  
49 Eight trials randomising 901 participants reported on quality of life,<sup>20;21;37;39;55;59;70;79</sup> observing that  
50  
51 participants allocated to exercise did not have significantly better quality of life (SMD 0.43; 95% CI -0.04 to  
52  
53 0.91; P=0.08). The I<sup>2</sup> was 89% showing substantial heterogeneity (Figure S8).  
54  
55  
56  
57  
58  
59  
60



### *Non-serious adverse events*

Non-serious adverse events were reported in only nine trials.<sup>20;21;38;55;57;59;64;66;67</sup> Five trials reported on musculoskeletal adverse events without conducting formal tests<sup>57;59;64;66;67</sup> and four trials reported on number of participants with high depression scores post-intervention compared to baseline assessment.<sup>20;21;64;67</sup> The RR for increased severity of depression post-intervention was 0.83 (95% CI 0.40 to 1.70; P=0.60; I<sup>2</sup> = 0.0%).

### *The effect of exercise on depression beyond the duration of the intervention*

Assessment of depression beyond the intervention was conducted in seven trials,<sup>20;37;39;51;59;62;80</sup> with a median duration between end of intervention and assessment of depression of 6 months (range 5 to 23.5 months). The SMD between the intervention group and the control group using a random effects analysis was -0.10 (95% CI -0.28 to 0.09; P=0.31; I<sup>2</sup> = 19.5%). The I<sup>2</sup> for this estimate was 19.5% suggesting low heterogeneity (See Figure S9).

Remission beyond the intervention was assessed in five trials,<sup>20;37-39;53</sup> and the relative risk of lack of remission was 0.95 (95% CI 0.82 to 1.11; P=0.53) with an I<sup>2</sup> of 0.0% (See Figure S10).

### *GRADE assessments*

The GRADE assessments are presented in Table 4, and quality of evidence for both primary and secondary outcomes was very low or low.

### Additional analysis

Four studies reported change in scores from baseline with corresponding SD's, and one study reported mean difference between groups post-intervention. Comparing the effect size of these five studies with the remaining did not explain part of the heterogeneity ( $p = 0.23$ ).

### Discussion

Thirty-one clinical trials allocating more than 2400 participants diagnosed with depression according to validated diagnostic instruments were included in the present systematic review. Pooled estimates suggested moderate antidepressant effect assessed both as a continuous outcome and as lack of remission. Due to risk of bias, inconsistency of effect estimates, and publication bias we have, however, very little confidence in these effect estimates. Subgroup analyses exploring reasons for the heterogeneity found that trials potentially having less risk of bias than other trials at high risk of bias had no effect of exercise on depression. Furthermore, duration of intervention and trial size were inversely associated with effect estimates. Exercise did not improve quality of life or depression or remission after the intervention. Serious adverse event or adverse events were reported inconsistently and only by a few trials not permitting firm conclusions regarding these outcomes.

### *Strengths and limitations*

The strengths of this systematic review are that it is based on the published protocol, a comprehensive search strategy, and the inclusion of patient centered outcomes such as quality of life as well as adverse events. Also, to avoid spurious finding from repeated testing, Trial Sequential Analysis and Bayes analysis were undertaken and these analyses did not suggest that the pooled estimates could be reduced to random errors for effect on depression severity or no remission. Neither Trial Sequential Analysis nor Bayes

1  
2  
3  
4 factor analysis are, however, able to wash of spurious effects induced by bias, or fraud or other  
5  
6 reasons.<sup>25;28;81-83</sup> Had we restricted the Trial Sequential Analysis to trials of potentially lower risk of bias, the  
7  
8 number of trials and participants would be limited and we had seen evidence far from crossing boundaries  
9  
10 for benefit, harms, or futility. The conclusions for serious adverse events and adverse events were  
11  
12 associated with wide confidence intervals due to lack of data and firm conclusions for these outcomes are  
13  
14 presently not available.  
15  
16  
17  
18  
19

20  
21 The number of trials with adequate allocation concealment was 39% in the current systematic review  
22  
23 compared to only 15.1% in trials assessing non-drug interventions for depression.<sup>84</sup> Blinded outcome  
24  
25 assessment was performed in 52% of the included trials compared to 44% in non-drug antidepressant trials  
26  
27 in general.<sup>84</sup> The incomplete outcome bias domain was adequate in 48% of our included trials compared to  
28  
29 32.9% of antidepressant non-drug trials in general.<sup>84</sup> Compared to non-drug trials assessing interventions  
30  
31 for participants with depression, the included exercise trials have more bias domains with low risk of bias.  
32  
33 However, all our included trials were at high risk of bias. Two trials had low risk of bias for all bias domains  
34  
35 except for blinding of participants and trial personnel, and four trials fulfilled our criteria for trials at  
36  
37 potentially less risk of bias than the rest of the trials with at risk of bias. Despite a search strategy including  
38  
39 bibliographical databases and trials from China and South-America, the vast majority of included trials were  
40  
41 conducted in north America and western Europe, which is comparable to the geographical distribution of  
42  
43 non-drug trials in general<sup>84</sup> limiting the applicability to other geographic regions.  
44  
45  
46  
47  
48  
49

50 All outcomes for the primary analysis reflect depression severity, however, the different psychometrics may  
51  
52 represent different aspects of depression not reflected in the pooled estimate. An in-depth discussion of  
53  
54 the included assessment scales is beyond the scope of this review, but in the current systematic review we  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 found no significant differences of effect estimates from trials using HAM-D<sub>17</sub> compared to trials using  
5  
6 other assessment scales (data not shown).  
7  
8  
9

### 10 11 *The effect of exercise on depression*

12  
13  
14 Our present results are similar to the latest Cochrane review by Cooney et al. (2013)<sup>23</sup> who found a  
15 moderate effect of exercise on depressive symptoms (-0.62 SMD) when including all trials and no effect  
16 when restricting the analysis to trials with less risk of bias (-0.18 SMD). The Cochrane review did find  
17 evidence of a small antidepressant effect beyond the intervention, which we could not confirm in our  
18 present systematic review. Bridle et al. (2012)<sup>12</sup> included 9 trials allocating old (> 60 years) participants with  
19 depression to exercise interventions versus control interventions. Restricting the analysis to four trials at  
20 lower risk of bias they found small to moderate effect estimates (SMD -0.34) in favour of exercise. The  
21 studies by Cooney et al.<sup>23</sup> and Bridle et al.<sup>12</sup> both included trials allocating participants with depressive  
22 symptoms and not necessarily diagnosed using a validated diagnostic system, potentially explaining the  
23 differences in the effect sizes. However, in our present systematic review the estimate for four trials at  
24 potential less risk of bias than the remaining trials was -0.11 SMD and in the Cooney study the effect  
25 estimate for eight trials with lower risk of bias was -0.18 SMD<sup>23</sup> compared to -0.34 in the study by Bridle et  
26 al.<sup>12</sup> Meta-analysis of randomised clinical trials assessing the effects of exercise for depression consistently  
27 finds positive effects, however, when restricting the analysis to trials with less risk of bias the pooled effect  
28 sizes becomes very small or negligible. Meta-analysis examining the effect of exercise beyond the  
29 intervention also finds no or small effects of exercise. In the process of interpretation of effect estimates in  
30 the current research field, it is important to recognise that effect estimates from trials with non-blinded  
31 outcome assessment are at high risk of bias as reported by Savovic et al.<sup>85</sup> Thirteen of 31 trials in the  
32 current systematic review did not use blinded outcome assessment. In contradiction to the current  
33 systematic review, a recent meta-analysis by Schuch et al.<sup>11</sup> concluded that “exercise has a large and  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 significant antidepressant effect in people with depression.....Our data strongly support the claim that  
5  
6 exercise is an evidence-based treatment for depression". This statement was based on a meta-analysis of  
7  
8 25 randomised clinical trials including participants with depression or depressive symptoms to exercise or  
9  
10 control conditions and excluding trials using any form of active control group. Surprisingly, the authors  
11  
12 found that adjusting for publication bias using the Trim and Fill procedure<sup>30</sup> the estimate *increased* from a  
13  
14 SMD of 0.98 to 1.11. The effect in SMD in included studies ranged from -0.23 to 4.56 representing  
15  
16 considerable heterogeneity.<sup>11</sup> The authors classified four trials as having lower risk of bias using the same  
17  
18 criteria as in our systematic review and 21 trials as having high risk of bias. This illustrates some of the  
19  
20 challenges in meta-analysis of exercise and depression: the large heterogeneity driven by small studies  
21  
22 inflating the effects of random-effects analysis,<sup>86</sup> the misconception that we can restrict our analysis to  
23  
24 statistics and not consider the evident effect of bias.<sup>22:85</sup> Compared to our previous review,<sup>10</sup> we now  
25  
26 included 31 trials including 2419 participants versus previously 13 trials and 687 participants. It may seem  
27  
28 as a paradox that this large increase in data has not provided us with a similar increase in certainty of  
29  
30 conclusions reflected by heterogeneity of trial results as well as our conclusions from the systematic  
31  
32 reviews. The increase in available data is, however, primarily provided by small trials at high risk of bias  
33  
34 introducing exaggerated effect estimates. In the current systematic review, we included four trials with 530  
35  
36 participants at lower risk of bias compared to three trials with 239 participants in our previous review,  
37  
38 reflecting that only a small part of the additional data comes from trials at lower risk of bias. The  
39  
40 continuous increase in data associated with high risk of bias will not provide patients, clinicians or  
41  
42 policymakers with adequate information and represents an unethical enrollment of trial participants and  
43  
44 waste of resources.<sup>87-93</sup> We therefore recommend that future systematic reviews and meta-analysis a priori  
45  
46 should have a primary outcome restricting effect analysis to larger trials with lower risk of bias and that any  
47  
48 recommendations regarding exercise interventions for participants with depression should be assessed  
49  
50 with the GRADE framework.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 The  $I^2$  of 82% and 71% for the primary outcomes indicate substantial evidence of heterogeneity of  
5  
6 intervention effects that is variation in effect estimates beyond chance. Part of this heterogeneity was  
7  
8 explained by bias and by trial size: trials at high risk of bias or small trials have very large effect estimates  
9  
10 compared to trials potentially at less bias risk compared to the remaining trials at high risk of bias or larger  
11  
12 trials. The funnel plots end Egger's test indicates publication bias, however, the association between trial  
13  
14 size and effect estimates could suggest that the asymmetry in the funnel plots are due to small study bias  
15  
16 rather than publication bias.<sup>94</sup> It could be argued that both the delivery of exercise as well as the actual  
17  
18 increase in fitness are fundamental to the assessment of the antidepressant effects of exercise, and in line  
19  
20 with our previous review we found duration of intervention inversely associated with effect size.<sup>10</sup>  
21  
22 Comparing different exercise intensities, we did find a small effect of high intensity exercise compared to  
23  
24 lower intensity exercise. However, assessing delivered exercise expressed as increase in maximal oxygen  
25  
26 uptake we could not reproduce this finding. Future trials need to pay more attention to the dose of the  
27  
28 intervention as well as compliance with intervention.<sup>95</sup> We suggest using maximal oxygen uptake or 1  
29  
30 repetition maximum as the gold standard to assess the received exercise. Several studies compare exercise  
31  
32 to control interventions rather than wait-list control to reduce the effect of non-specific effects, e.g., the  
33  
34 DEMO trials and Mather et al.<sup>20;21;51</sup> Also, it could be speculated that the effect of exercise would be harder  
35  
36 to detect if participants also received medical treatment in addition. The current systematic review could  
37  
38 not confirm that the type of control condition explained heterogeneity. The discussion of control group is  
39  
40 important in non-drug trials: choosing a waitlist control group the results potentially reflects non-specific  
41  
42 effects, choosing an active control group (e.g., relaxation exercise) the trial is potentially a comparison  
43  
44 between to active treatments. However, in the current systematic review we found no evidence that trials  
45  
46 using an attention control group or exercise as add-on to pharmacotherapy had significantly different effect  
47  
48 estimates compared to other trials.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 Our systematic review did not find indications of a positive effect on quality of life in participants with  
5 depression allocated to exercise interventions, which is in concordance with the review by Cooney et al.<sup>23</sup>  
6  
7  
8 Only 3/31 trials reported on serious adverse events, and found no significant risk of death or suicide  
9  
10 attempt. No indication of increased severity of depression or other adverse events in participants allocated  
11 to exercise could be detected. However, data on adverse events was reported sporadically in a minority of  
12 trials and currently it is not possible to conclude on the risk of serious adverse events or adverse event from  
13 exercise interventions in participants with depression.  
14  
15  
16  
17  
18  
19  
20  
21  
22

### 23 *Conclusions*

24  
25 We have little confidence in the pooled effect estimates, especially because trials with less than high risk of  
26 bias produced significantly lower effect estimates, suggesting that exercise interventions only produce  
27 small or negligible antidepressant effects, depending on how much of the effect is caused by bias and how  
28 much is caused by the intervention. There was no effect of exercise on quality of life or depression beyond  
29 the intervention itself. There is currently no evidence in favour of exercise for patients with depression with  
30 a view to ameliorate depressive symptoms and at we do not recommend that exercise is prescribed to  
31 relieve depressive symptoms. Our systematic review did not evaluate possible beneficial effects of exercise  
32 on, e.g., metabolism or cardiovascular fitness,<sup>21,96</sup> and it is possible that exercise may have beneficial effects  
33 on these factors in patients diagnosed with depression.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

### 49 *Future perspectives*

50  
51 Despite the large number of published trials, further trials with more robust methodology seem still  
52 required to establish progress in this field. Also, additional trials from outside North-America and Europe  
53 may be required for results to be valid for patients in Asia, Africa, and South-America. To further elaborate  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 on the current findings, we recommend that future trials must include blinded outcome assessors and  
5  
6 outcomes assessing quality of life, metabolic effects, and long-term effects beyond the intervention. It is  
7  
8 also important that future trials systematically collect and report data on death, suicide events,  
9  
10 musculoskeletal injuries and other potential adverse effects in both the intervention group as well as in the  
11  
12 control group. Moreover, future trials ought to be designed according to the SPIRIT guidelines and reported  
13  
14 according to the CONSORT guidelines<sup>97,98</sup> and transparently report deidentified individual participant data  
15  
16 enabling individual participant data meta-analyses.<sup>99</sup>  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## Funding

No funding was obtained for this systematic review.

## Competing interests

JK, CG, and MN have previously published two trials and a meta-analysis on this topic, which could introduce an academic bias in the current systematic review. We asked new authors (HS and CH) to be involved in the preparation of the protocol, trial selection and bias assessment. No support from any organisation was received for the submitted work; no financial relationship with any organisations that might have an interest in the submitted work in the previous three years, and apart from the above no other relationship or activities that could appear to have influenced the submitted work.

## Acknowledgement

We appreciate the help from Youling He for help with the Chinese Wanfang bibliographical database and translation of Chinese papers. We thank Janus C Jakobsen for assistance with the calculation of Bayes factor.

## Contributors

JK conceived the project, collected data, did the statistical analysis, analysed the data, drafted and revised the manuscript. He is guarantor. CH collected the data, analysed the data and revised the

1  
2  
3  
4 manuscript. HS conceived the project, collected data, analysed the data, and revised the  
5  
6 manuscript. CG conceived the project, analysed the data and revised the manuscript. MN  
7  
8 conceived the project, analysed the data, and revised the manuscript.  
9  
10

#### 11 12 13 14 15 **Data sharing statement**

16  
17  
18 All data used in this study are available in Figures and Tables. No other data were used.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Reference List

- 1  
2  
3  
4  
5  
6  
7  
8 (1) Lepine JP, Gastpar M, Mendelwicz J, Tylee A. Depression in the community: The first pan-  
9 european study DEPRES (Depression Research in European Society). *Int Clin Psychopharmacol*  
10 1997;12:19-29.
- 11  
12 (2) Pedersen CB, Mors O, Bertelsen A, *et al*. A comprehensive nationwide study of the incidence rate  
13 and lifetime risk for treated mental disorders. *JAMA Psychiatry* 2014;71(5):573-81.
- 14  
15 (3) Ustun TB, Ayuso-Mateos JL, Chatterji S, *et al*. Global burden of depressive disorders in the year  
16 2000. *Br J Psychiatry* 2004;184:386-92.
- 17  
18 (4) Kirsch I, Deacon B, Huedo-Medina T, *et al*. Initial severity and antidepressant benefits: A meta-  
19 analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5[2], e45.  
20 doi:10.1371/journal.pmed.0050045.
- 21  
22 (5) Turner EH, Matthews AM, Linardatos E, *et al*. Selective Publication of Antidepressant Trials and  
23 Its Influence on Apparent Efficacy. *N Engl J Med* 2008; 358(3):252-60.
- 24  
25 (6) Jakobsen JC, Hansen JL, Storebo OJ, *et al*. The effects of cognitive therapy versus 'no intervention'  
26 for major depressive disorder. *PLoS One* 2011;6(12):e28299.
- 27  
28 (7) Jakobsen JC, Hansen JL, Simonsen E, Gluud C. The effect of interpersonal psychotherapy and  
29 other psychodynamic therapies versus 'treatment as usual' in patients with major depressive  
30 disorder. *PLoS One* 2011; 6(4):e19044.
- 31  
32 (8) Jakobsen JC, Hansen JL, Simonsen E, Gluud C. The effect of adding psychodynamic therapy to  
33 antidepressants in patients with major depressive disorder. A systematic review of randomized  
34 clinical trials with meta-analyses and trial sequential analyses. *J Affect Disord* 2012;137(1-3):4-14.
- 35  
36 (9) Cipriani A, Furukawa TA, Salanti G, *et al*. Comparative efficacy and acceptability of 12 new-  
37 generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009;373[9665]:746-58.
- 38  
39 (10) Krogh J, Nordentoft M, Sterne J, Lawlor D. The effect of exercise in clinically depressed adults:  
40 Systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry*  
41 2011;72(4):529-38.
- 42  
43 (11) Schuch FB, Vancampfort D, Richards J, *et al*. Exercise as a treatment for depression: A meta-  
44 analysis adjusting for publication bias. *J Psychiatr Res* 2016;77:42-51.
- 45  
46 (12) Bridle C, Spanjers K, Patel S, *et al*. Effect of exercise on depression severity in older people:  
47 systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry* 2012;  
48 201(3):180-185.
- 49  
50 (13) Herring MP, Puetz TW, O'Connor PJ, Dishman RK. Effect of exercise training on depressive  
51 symptoms among patients with a chronic illness: a systematic review and meta-analysis of  
52 randomized controlled trials. *Arch Intern Med* 2012; 172(2):101-111.
- 53  
54  
55  
56  
57  
58  
59  
60

- 1
- 2
- 3
- 4 (14) World Health Organisation. International Statistical Classification of Diseases, 10th Revision (ICD-10). Geneva, Switzerland 1992.
- 5
- 6
- 7 (15) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. Washington DC 1994.
- 8
- 9
- 10 (16) Beck AT, Steer RA, Brown GK. BDI-II manual. 2 ed. New York: Psychological Corporation, 1996.
- 11
- 12 (17) Krogh J, Speyer H, Gluud C, Nordentoft M. Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential analysis. *Syst Rev* 2015; 4(1):40.
- 13
- 14 (18) Spitzer RL, Endicott J, Williams JB. Research diagnostic criteria. *Arch Gen Psychiatry* 1979; 36(12):1381-1383.
- 15
- 16 (19) ICH-secretariat. International Conference on Harmonisation (ICH) Harmonised Tripartite Guideline (Step 4). Guideline for Good Clinical Practice. Geneva 1996.
- 17
- 18 (20) Krogh J, Saltin B, Gluud C, Nordentoft M. 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;70[6]: 790-00.
- 19
- 20 (21) Krogh J, Videbech P, Thomsen C, *et al.* DEMO-II trial. Aerobic Exercise Versus Stretching Exercise in Patients With Major Depression - A Randomized Clinical Trial. *PLoS One* 2012;7(10): e48316. doi:10.1371/journal.pone.0048316
- 21
- 22 (22) Higgins JPT. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org) The Cochrane Collaboration, 2009.
- 23
- 24 (23) Cooney GM, Dwan K, Greig CA, *et al.* Exercise for depression. *Cochrane Database Syst Rev* 2013; 9:CD004366.
- 25
- 26 (24) Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ* 2003; 327:557-60.
- 27
- 28 (25) Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008; 61(1):64-75.
- 29
- 30 (26) Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008; 61(8):763-69.
- 31
- 32 (27) Thorlund K, Walter SD, Johnston BC, *et al.* Pooling health-related quality of life outcomes in meta-analysis-a tutorial and review of methods for enhancing interpretability. *Res Synth Methods* 2011; 2(3):188-03.
- 33
- 34 (28) Jakobsen JC, Wetterslev J, Winkel P, *et al.* Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. *BMC Med Res Methodol* 2014; 14:120.
- 35
- 36 (29) Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 1999; 319(7211):670-74.
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1  
2  
3  
4 (30) Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for  
5 publication bias in meta-analysis. *Biometrics* 2000; 56(2):455-63.  
6  
7 (31) Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of  
8 evidence and strength of recommendations. *BMJ* 2008; 336(7650):924-26.  
9  
10 (32) Andrews J, Guyatt G, Oxman AD, *et al.* GRADE guidelines: 14. Going from evidence to  
11 recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;  
12 66(7):719-25.  
13  
14 (33) Zheng YP, Lin KM, Zhao JP, *et al.* Comparative study of diagnostic systems: Chinese Classification  
15 of Mental Disorders-Second Edition versus DSM-III-R. *Compr Psychiatry* 1994; 35(6):441-49.  
16  
17 (34) Babyak MA, Blumenthal JA, Herman S, *et al.* Exercise treatment for major depression:  
18 Maintenance of therapeutic benefit at 10 months. *Psychosom Med* 2000; 62:633-38.  
19  
20 (35) Baxter H, Winder R, Chalder M, *et al.* Physical activity as a treatment for depression: the TREAD  
21 randomised trial protocol. *Trials* 2010; 11:105.  
22  
23 (36) Benson M, Blumenthal J, Babyak M, *et al.* Exercise fails to improve neurocognition in depressed  
24 middle-aged and older adults. *Med Sci Sports Exerc* 2008;40[7]:1344-352.  
25  
26 (37) Blumenthal J, Babyak MA, Moore KA, *et al.* Effects of exercise training on older patients with  
27 major depression. *Arch Int Med* 1999;159:2349-356.  
28  
29 (38) Blumenthal J, Babyak M, Doraiswamy P, *et al.* Exercise and pharmacotherapy in the treatment of  
30 major depressive disorder. *Psychosom Med* 2007;69:587-96.  
31  
32 (39) Chalder M, Wiles NJ, Campbell J, *et al.* Facilitated physical activity as a treatment for depressed  
33 adults: randomised controlled trial. *BMJ* 2012; 344:e2758.  
34  
35 (40) Chalder M, Wiles NJ, Campbell J, *et al.* A pragmatic randomised controlled trial to evaluate the  
36 cost-effectiveness of a physical activity intervention as a treatment for depression: the treating  
37 depression with physical activity (TREAD) trial. *Health Technol Assess* 2012; 16(10):1-iv.  
38  
39 (41) Combs K, Smith PJ, Sherwood A, *et al.* Impact of sleep complaints and depression outcomes  
40 among participants in the standard medical intervention and long-term exercise study of exercise  
41 and pharmacotherapy for depression. *J Nerv Ment Dis* 2014;202(2):167-71.  
42  
43 (42) Doyne EJ, Ossip-Klein DJ, Bowman ED, *et al.* Running versus weight lifting in the treatment of  
44 depression. *J Consult Clin Psychol* 1987;55(5):748-54.  
45  
46 (43) Dunn AL, Trivedi MH, Kampert JB, *et al.* The DOSE study: a clinical trial to examine efficacy and  
47 dose response of exercise as treatment for depression. *Control Clin Trials* 2002;23(5):584-03.  
48  
49 (44) Dunn AL, Trivedi MH, Kaplan K, *et al.* Exercise treatment for depression. Efficacy and dose  
50 response. *Am J Prev Med* 2005;28:1-8.  
51  
52 (45) Herman S, Blumenthal JA, Babyak M, *et al.* Exercise therapy for depression in middle-aged and  
53 older adults: predictors of early dropout and treatment failure. *Health psychology* 2002;21:553-  
54 63.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (46) Ho CWH, Chan SC, Wong JS, *et al.* Effect of aerobic exercise training on chinese population with  
5 mild to moderate depression in Hong Kong. *Rehabil Res Pract* 2014;2014:627376  
6  
7 (47) Hoffman BM, Babyak MA, Craighead WE, *et al.* Exercise and pharmacotherapy in patients with  
8 major depression: one-year follow-up of the SMILE study. *Psychosom Med* 2011;73(2):127-33.  
9  
10 (48) Kerling A, Tegtbur U, Gutzlaff E, *et al.* Effects of adjunctive exercise on physiological and  
11 psychological parameters in depression: A randomized pilot trial. *J Affect Disord* 2015;177:1-6.  
12  
13 (49) Khatri P, Blumenthal J, Babyak M, *et al.* Effects of exercise training on cognitive functioning  
14 among depressed older men and women. *J Aging Phys Act* 2001;9:43-57  
15  
16 (50) Krogh J, Petersen L, Timmermann M, *et al.* Design paper: The DEMO trial: A randomized, parallel-  
17 group, observer-blinded clinical trail of aerobic versus non-aerobic versus relaxation training for  
18 patients with light to moderate depression. *Contemp Clin Trials* 2007;28:79-89.  
19  
20 (51) Mather AS, Rodriguez C, Guthrie MF, *et al.* Effects of exercise on depressive symptoms in older  
21 adults with poorly responsive depressive disorder: randomised controlled trial. *Br J Psychiatry*  
22 2002;180:411-15.  
23  
24 (52) Mota-Pereira J, Silverio J, Carvalho S, *et al.* Moderate exercise improves depression parameters in  
25 treatment-resistant patients with major depressive disorder. *J Psychiatr Res* 2011; 45(8):1005-  
26 011.  
27  
28 (53) Pfaff JJ, Alfonso H, Newton RU, *et al.* ACTIVEDEP: a randomised, controlled trial of a home-based  
29 exercise intervention to alleviate depression in middle-aged and older adults. *Br J Sports Med*  
30 2014;48(3):226-32.  
31  
32 (54) Pilu A, Sorba M, Hardoy M, *et al.* Efficacy of physical activity in the adjunctive treatment of major  
33 depressive disorders: preliminary results. *Clin Pract Epidemio Ment Health* 2007;3:8.  
34  
35 (55) Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Fleck MP. Exercise and severe depression:  
36 preliminary results of an add-on study. *J Affect Disord* 2011;133(3):615-18.  
37  
38 (56) Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of the effect of exercise on  
39 sleep. *Sleep* 1997; 20(2):95-01.  
40  
41 (57) Singh NA, Stavrinou TM, Scarbek Y, *et al.* A randomized controlled trial of high versus low  
42 intensity weight training versus general practitioner care for clinical depression in older adults. *J*  
43 *Gerontol A Biol Sci Med Sci* 2005; 60(6):768-776.  
44  
45 (58) Singh NA, Clements KM, Singh MA. The efficacy of exercise as a long-term antidepressant in  
46 elderly subjects: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* 2001; 56(8):M497-  
47 M504.  
48  
49 (59) Singh N, Clements K, Fiatarone M. A randomized controlled trial of progressive resistance training  
50 in depressed elders. *J Gerontol* 1997; 52:M27-M35.  
51  
52 (60) Toni G, Belvederi MM, Zanetidou S, *et al.* Physical exercise for late life Major Depression: The  
53 SEEDS study. *Eur J Prev Cardiol* 2014; 21:S107.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (61) Veale D, Le FK, Pantelis C, *et al.* Aerobic exercise in the adjunctive treatment of depression: a  
5 randomized controlled trial. *J R Soc Med* 1992; 85(9):541-544.  
6  
7 (62) Vieira JLL, Porcu M, da Rocha PGM. The regular physical exercise's practice as complementary  
8 therapy to the treatment of depressed women. *J Bras Psiquiatr* 2007;56(1):23-8.  
9  
10 (63) Epstein D. Reno: University of Nevada: Aerobic activity versus group cognitive therapy: An  
11 evaluative study of contrasting interventions for the alleviation of clinical depression.  
12 (*dissertation*) 1986.  
13  
14 (64) Danielsson L, Papoulias I, Petersson E, *et al.* Exercise or basic body awareness therapy as add-on  
15 treatment for major depression: A controlled study. *J Affect Disord* 2014;168:98-06.  
16  
17 (65) Carneiro L, Fonseca A, Vieira-Coelho M, *et al.* Effects of structured exercise and pharmacotherapy  
18 vs pharmacotherapy for adults with depressive symptoms: A randomized clinical trial. *J Psychiatr*  
19 *Res* 2015;(71):48-55.  
20  
21 (66) Legrand FD, Neff EM. Efficacy of exercise as an adjunct treatment for clinically depressed  
22 inpatients during the initial stages of antidepressant pharmacotherapy: An open randomized  
23 controlled trial. *J Affect Disord* 2016;191:139-44.  
24  
25 (67) Doose M, Ziegenbein M, Hoos O, *et al.* Self-selected intensity exercise in the treatment of major  
26 depression: A pragmatic RCT. *Int J Psychiatry Clin Pract* 2015;19(4):266-75.  
27  
28 (68) Salehi I, Hosseini SM, Haghighi M, *et al.* Electroconvulsive therapy (ECT) and aerobic exercise  
29 training (AET) increased plasma BDNF and ameliorated depressive symptoms in patients suffering  
30 from major depressive disorder. *J Psychiatr Res* 2016;76:1-8.  
31  
32 (69) Huipeng F, Xiaohui Z. Curative effect and observation of exercise on patients with depression.  
33 *Chin Med Pharmacy* 2013;3:187-88.  
34  
35 (70) Jinchun W, Bo Y, Rong X, *et al.* Effect of exercise therapy in improving social function and quality  
36 of life in patients with depression. *Nurs J Chin PLA* 2015; 32(21):24.  
37  
38 (71) Guifeng S. The role of setting - up exercise on rehabilitation in patients with depressive episode.  
39 *Chin J Health Psychol* 2015;23:184-186.  
40  
41 (72) Fang D, Min R, Ai-qin W. The effect of aerobic exercise with different intensity and frequency on  
42 depressive symptoms. *Chin J Behav Med & Brain Sci* 2013;22:830-32.  
43  
44 (73) Martinsen EW, Medhus A, Sandvik L. Effects of aerobic exercise on depression: A controlled  
45 study. *BMJ* 1985; 291:109.  
46  
47 (74) Veale D, Fevre K, Pantalis C, *et al.* Aerobic exercise in the adjunctive treatment of depression: A  
48 randomised controlled trial. *J R Soc Med* 1992;85:541-44.  
49  
50 (75) Ossip-Klein DJ, Doyne EJ, Bowman ED, *et al.* Effects of running or weight lifting on self-concept in  
51 clinically depressed women. *J Consult Clin Psychol* 1989;57(1):158-61.  
52  
53 (76) Blumenthal JA, Babyak MA, Moore KA, *et al.* Effects of exercise training on older patients with  
54 major depression. *Arch Intern Med* 1999; 159(19):2349-356.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (77) Belvederi MM, Amore M, Menchetti M, *et al.* Physical exercise for late-life major depression. *Br J Psychiatry* 2015; 207(3):235-42.
- 5  
6  
7 (78) 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.
- 8  
9  
10  
11 (79) Carta MG, Hardoy MC, Pilu A, *et al.* Improving physical quality of life with group physical activity in the adjunctive treatment of major depressive disorder. *Clin Pract Epidemiol Ment Health* 2008; 4:1.doi: 10.1186/1745-0179-4-1.
- 12  
13  
14  
15 (80) Klein MH, Greist JH, Gurman AS, *et al.* A comparative outcome study of group psychotherapy vs. exercise treatments for depression. *Int J Ment Health* 1985;13:148-77.
- 16  
17  
18  
19 (81) Roberts I, Ker K, Edwards P, *et al.* The knowledge system underpinning healthcare is not fit for purpose and must change. *BMJ* 2015; 350:h2463.
- 20  
21  
22 (82) Gluud C, Jakobsen J, Imberger G, *et al.* Re: The knowledge system underpinning healthcare is not fit for purpose and must change - Responses to the opposing viewpoints of Roberts and colleagues and Tovey and colleagues. *BMJ* 2015; 350(h2463).
- 23  
24  
25  
26 (83) Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for Trial Sequential Analysis (TSA). *Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen, Denmark P1-115, Available from <https://www.ctu.dk/tsa> 2011.*
- 27  
28  
29  
30 (84) Krogh J, Hjorthoj CR, Jakobsen JC, *et al.* DEPEROR: Risks of systematic errors in drug and non-drug randomized clinical trials assessing intervention effects in patients with unipolar depression. *J Affect Disord* 2015; 179:121-27.
- 31  
32  
33  
34 (85) Savovic J, Jones HE, Altman DG, *et al.* Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012; 157(6):429-38.
- 35  
36  
37  
38 (86) Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol* 1999; 150(5):469-75.
- 39  
40  
41  
42 (87) Ioannidis JP, Greenland S, Hlatky MA, *et al.* Increasing value and reducing waste in research design, conduct, and analysis. *Lancet* 2014; 383(9912):166-75.
- 43  
44  
45  
46 (88) Chan AW, Song F, Vickers A, *et al.* Increasing value and reducing waste: addressing inaccessible research. *Lancet* 2014; 383(9913):257-66.
- 47  
48  
49 (89) Moher D, Glasziou P, Chalmers I, *et al.* Increasing value and reducing waste in biomedical research: who's listening? *Lancet* 2016; 387(10027):1573-586.
- 50  
51  
52 (90) Glasziou P, Altman DG, Bossuyt P, *et al.* Reducing waste from incomplete or unusable reports of biomedical research. *Lancet* 2014; 383(9913):267-76.
- 53  
54  
55 (91) Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM *et al.* How to increase value and reduce waste when research priorities are set. *Lancet* 2014; 383(9912):156-65.
- 56  
57  
58  
59  
60



- 1  
2  
3  
4 (92) Al-Shahi SR, Beller E, Kagan J, *et al.* Increasing value and reducing waste in biomedical research  
5 regulation and management. *Lancet* 2014;383(9912):176-85.  
6  
7 (93) Garattini S, Jakobsen JC, Wetterslev J, *et al.* Evidence-based clinical practice: Overview of threats  
8 to the validity of evidence and how to minimise them. *Eur J Intern Med* 2016;32:13-21.  
9  
10 (94) Sterne JA, Sutton AJ, Ioannidis JP, *et al.* Recommendations for examining and interpreting funnel  
11 plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; 343:d4002.  
12  
13 (95) Slade SC, Dionne CE, Underwood M, *et al.* Consensus on Exercise Reporting Template (CERT):  
14 Explanation and Elaboration Statement. *Br J Sports Med* 2016.  
15  
16 (96) Stubbs B, Rosenbaum S, Vancampfort D, *et al.* Exercise improves cardiorespiratory fitness in  
17 people with depression: A meta-analysis of randomized control trials. *J Affect Disord* 2016;  
18 190:249-53.  
19  
20 (97) Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting  
21 parallel group randomized trials. *Ann Intern Med* 2010; 152(11):726-32.  
22  
23 (98) Chan AW, Tetzlaff JM, Gotzsche PC, *et al.* SPIRIT 2013 explanation and elaboration: guidance for  
24 protocols of clinical trials. *BMJ* 2013; 346:e7586.  
25  
26 (99) Skoog M, Saarimäki J, Gluud C, *et al.* Transparency and registration in clinical research in the  
27 Nordic countries. *Nordic Trial Alliance, NordForsk*; 2015 1-108.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 **Figure 1.** Effect of exercise on depression severity in patients diagnosed with depression  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

**Table 1.** Characteristics of trials assessing exercise for patients diagnosed with depression

Author, first Country of origin	Participants	Severity of depression at baseline	N at baseline (included in trial efficacy analysis)	Type of intervention	Frequency	Duration
Klein 1985 USA	Outpatients Mean age: 30 (SD 7) 72% female	SCL-D: 2.4 (SD 1)	50 (22)	<i>Aerobic exercise:</i> Supervised individual running. <i>Control group:</i> Supervised meditation in groups	2 sessions per week  Control group: 1 session per week	12 weeks
Martinsen 1985 Norway	Inpatients Mean age: 40 (range 17-60) Distribution of sex not reported	BDI: 28.0 (SD 9)	49(43)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Occupational therapy.	3 sessions per week  Control group: 3 sessions per week	9 weeks
Epstein 1986 USA	Outpatients Mean age: 39 (range 24 to 60) (NR) % female	BDI: 23.4 (SD 7)	21 (17)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Waitlist control.	3 sessions per week	8 weeks
Doyne 1987 USA	Outpatients Mean age: 29 (SD 4) 100 % female	HAM-D <sub>17</sub> : 13.0 (SD 7)	52 (25)	<i>Aerobic exercise OR weightlifting:</i> Supervised individual exercise. <i>Control group:</i> Waiting list.	4 sessions per week	8 weeks
Veale 1992 UK	Outpatients Mean age: 35 (range 19-58) 64% female	BDI: 24.5 (SD 6)	83 (65)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Standard treatment from psychiatric services.	3 sessions per week	12 weeks
Singh 1997 USA	Outpatients Recruited from a register of volunteers Mean age: 71 (SD 1)	BDI: 19.9 (SD 2.3)	32 (32)	<i>Progressive resistance training:</i> Supervised group exercise. <i>Control group:</i> Attended seminars on health.	3 sessions per week  Control group: 2 sessions per week	10 weeks
Blumenthal 1999 USA	Outpatients Mean age: 57 (SD 7) 71.8% female	HAM-D <sub>17</sub> : Not reported	103 (103)	<i>Aerobic exercise:</i> Supervised exercise plus antidepressant medication (sertraline). <i>Control group:</i> Antidepressant medication (sertraline).	3 sessions per week	16 weeks
Mather 2002 UK	Outpatients Treatment resistant Mean age: 65 (range 53-91) 69% female	HAM-D <sub>17</sub> : 17.1 (SD 6)	86 (85)	<i>Mixed aerobic and non-aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Attended health seminars.	2 sessions per week  Control group: 2 seminars per week	10 weeks
Dunn 2005 USA	Outpatients Mean age: 36 (SD 6)	HAM-D <sub>17</sub> : 19.4 (SD 2)	80 (80)	<i>Aerobic exercise:</i> Individually supervised	Group (1) and (2): 3 sessions	12 weeks

		75% female			exercise with (1) low energy expenditure (EE) OR (2) high EE OR (3) low EE OR (4) high EE. <i>Control group:</i> Flexibility exercise.	per week Group (3) and (4): 5 sessions per week Control group: 3 sessions per week	
Singh 2005 Australia	Outpatients Mean age: 69 (SD 6) 55% female	HAM-D <sub>17</sub> : 18.9 (SD 4.2)	60 (54)	<i>Progressive resistance training (PRT):</i> (1) Low intensity PRT OR (2) high intensity PRT. <i>Control group:</i> Standard GP care.	Group (1) and (2): 3 sessions per week	8 weeks	
Pilu 2007 Italy	Outpatients Treatment resistant Age between 40 and 60 100% female	HAM-D <sub>17</sub> : 19.7 (SD 6)	30 (30)	<i>Resistance exercise:</i> Supervised group sessions. <i>Control group:</i> Standard treatment.	2 sessions per week	32 weeks	
Viera 2007 Brazil	Outpatients Mean age 43.66 (SD NR) 100% female	HAM-D <sub>21</sub> : 31.9 (SD 3)	18 (18)	<i>Aerobic exercise:</i> Supervised water aerobics. <i>Control group:</i> Standard GP care.	2 sessions per week	12 weeks	
Blumenthal 2007 USA	Outpatients Mean age: 52 (SD 8) 75.8% female	HAM-D <sub>17</sub> : 16.7 (SD 4)	153 (153)	<i>Aerobic exercise:</i> (1) Supervised group exercise OR (2) home-based exercise. <i>Control group:</i> Placebo medication.	(1) and (2): 3 sessions per week	16 weeks	
Krogh 2009 Denmark	Outpatients Mean age: 39 (SD 9) 74% female	HAM-D <sub>17</sub> : 17.8 (SD 4)	165 (165)	<i>Exercise:</i> (1) Aerobic supervised group exercise OR (2) supervised group resistance training <i>Control group:</i> relaxation and stretching exercise.	(1) and (2): 2 sessions per week  Control group: 2 sessions per week	16 weeks	
Mota-Pereira 2011 Portugal	Outpatients Treatment resistant Mean age: 47.5 (SD 3) 65.5% female	HAM-D <sub>17</sub> : 17.1 (SD 3)	33 (29)	<i>Aerobic exercise:</i> Homebased exercise + supervised. <i>Control group:</i> Attention control.	4 home-based sessions/week. 1 supervised session/week Control group: 1 supervised session/week	12 weeks	
Krogh 2012 Denmark	Outpatients Mean age: 42 (SD 11) 67% female	HAM-D <sub>17</sub> : 18.9 (SD 4)	115 (115)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Supervised stretching exercise in groups.	3 sessions per week  Control group: 3 sessions per week	12 weeks	
Chalder 2012 UK	Outpatients Mean age: 40 (SD 13) 66% female	BDI: 32.1 (SD 9)	361 (361)	<i>Exercise:</i> Participants received individually tailored support and encouragement to engage in physical activity. <i>Control group:</i> Standard GP care.	Individual	16 weeks	

1						
2						
3						
4	Fang 2013	Inpatients	HAM-D <sub>24</sub> :	90 (90)	<i>Aerobic exercise:</i>	Group 1 and 2
5	China	Mean age: 44 (SD	29.2 (SD 5)		Group 1 and 2 had 3 and 5	6 weeks
6		14)			supervised group	sessions per
7		66.9% female			exercise, high	week,
8					intensity.	respectively
9					<i>Control group:</i>	Control group:
10					15 min stretching	3 sessions per
11	Huipeng 2013	Inpatients	HAM-D <sub>17</sub> :	68 (68)	<i>Aerobic exercise:</i>	5 sessions per
12	China	Mean age: 30 (SD 5)	28 (SD 5)		Jogging	week
13		100% female			<i>Control group:</i>	
14					Standard treatment	
15	Cassandra 2014	Inpatients	MADRS:	52 (52)	<i>Aerobic exercise:</i>	5 sessions per
16	Honkong	Mean age: 46 (SD	19 (10)		Supervised exercise.	week
17		12)			<i>Control group:</i>	
18		67.3% female			10 min stretching.	
19	Danielsson 2014	Outpatients	MADRS:	42 (42)	<i>Mixed aerobic and</i>	2 sessions per
20	Sweden	Mean age: 45 (SD	24.0 (SD 5)		<i>non-aerobic exercise:</i>	week
21		13)			First two weeks	
22		76% female			individual supervised	
23					exercise then	
24					supervised group	
25					exercise.	
26					<i>Control group:</i> One	
27					session with advice on	
28	Pfaff 2014	Outpatients	MADRS:	200 (200)	<i>Resistance exercise:</i>	3 sessions per
29	Australia	Mean age: 61 (SD 8)	21.3 (SD		Supervised home-	week
30		63% female	NR)		based exercise	
31					<i>Control group:</i>	
32					Standard GP care	
33	Guifeng 2015	Inpatients	HAM-D <sub>24</sub> :	70 (70)	<i>Aerobic exercise:</i>	5 sessions per
34	China	Mean age: 33 (SD	25.9 (SD 4)		Supervised group	week
35		14)			exercise	
36		70% female			<i>Control group:</i>	
37					Standard treatment	
38	Junchin 2015	Inpatients	HAM-D <sub>24</sub> :	70 (70)	<i>Aerobic exercise:</i>	5 sessions per
39	China	Mean age: 28 (SD 7)	25.8 (SD 3)		Supervised aerobic	week
40		61% female			exercise of the	
41					patients own choice	
42					<i>Control group:</i>	
43					Standard treatment	
44	Schuch 2015	Inpatients	HAM-D <sub>17</sub> :	50 (50)	<i>Aerobic exercise:</i>	3 sessions per
45	Brazil	Mean age: 40 (SD	26.7 (SD 2)		Supervised individual	week
46		11)			exercise.	
47		74% female			<i>Control group:</i>	
48					Standard treatment.	
49	Kerling 2015	Inpatients	MADRS:	42 (42)	<i>Aerobic exercise:</i>	3 sessions per
50	Germany	Mean age: 43 (SD	24.0 (SD 9)		Supervised exercise.	week
51		10)			<i>Control group:</i>	
52					Standard treatment.	
53	Belvederi 2015	Outpatients	HAM-D <sub>17</sub> :	121 (121)	<i>Aerobic exercise:</i>	3 sessions per
54	Italy	Mean age: 75 (SD 6)	20.1 (SD 3)		(1) Sertraline +	week
55		71% female			supervised non-	
56					progressive exercise	
57					OR (2) sertraline +	
58					supervised	
59					progressive aerobic	
60						

				exercise. <i>Control group:</i> Sertraline.		
Carneiro 2015 Portugal	Outpatients Mean age: 50.16 (SD 12) 100% female	BDI: 48.8 (SD 10)	26 (19)	<i>Aerobic exercise:</i> Supervised exercise <i>Control group:</i> Standard treatment	3 sessions per week	16 weeks
Doose 2015 Germany	Outpatients Mean age: 47.9 (SD 10.5) 63% female	HAM-D <sub>17</sub> : 14.2 (SD 3)	46 (46)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Standard treatment	3 sessions per week	8 weeks
Salehi 2016 Iran	Inpatients Mean age: 30.0 (SD 6) 35% female	HAM-D <sub>21</sub> : 43.4 (SD 8)	40 (40)	<i>Aerobic exercise + ECT:</i> Supervised aerobic exercise <i>Control group:</i> ECT	3 sessions per weeks  Control group 3 ECTs per week	4 weeks
Legrand 2016 France	Inpatients Mean age: 46.9 (SD 13) 67% female	BDI: 36.0 (SD 6)	24 (24)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Standard treatment	10 sessions in 10 consecutive days	10 days

SCL-D: Symptom Check List, depression subscale; HAM-D<sub>17</sub>: Hamilton Depression Scale, 17 items; BDI: Beck's Depression Inventory; SD: Standard deviation; ECT: Electroconvulsive therapy

**Table 2.** Risk of bias in trials assessing exercise for patients diagnosed with depression

Author, Year of publication	Sequence generation	Allocation concealment	Blinding of participants and trial personnel assessors	Blinding of outcome assessors	Incomplete outcome data	Selective outcome reporting	For profit bias	Other bias	Comment on 'Other bias'
Klein 1985	Unclear	Unclear	High	High	High	Low	Low	Low	
Martinsen 1985	Unclear	Unclear	High	High	High	Low	High	Low	
Epstein 1986	Unclear	Unclear	High	High	High	Low	Unclear	High	Baseline difference
Doynes 1987	Unclear	Unclear	High	Low	High	Low	Unclear	High	Baseline difference
Veale 1992	Unclear	Unclear	High	High	High	Low	Low	High	Baseline difference
Singh 1997	Low	Unclear	High	Low	Low	Low	Low	High	Baseline difference
Blumenthal 1999	Unclear	Unclear	High	Low	High	Low	High	Low	
Mather 2002	Low	Low	High	Low	High	Low	Low	Low	
Dunn 2005	Low	Low	High	Low	High	High	High	Low	
Singh 2005	Low	Low	High	Low	High	Low	Unclear	Low	
Pilu 2007	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Viera 2007	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Blumenthal 2007	Low	Low	High	Low	High	High	Low	Low	
Krogh 2009	Low	Low	High	Low	Low <sup>1</sup>	High	High	High	Baseline difference
Mota-Pereira 2011	Unclear	Unclear	High	Low	High	Low	High	High	Baseline difference
Krogh 2012	Low	Low	High	Low	Low	Low	Low	Low	
Chalder 2012	Low	Low	High	High	Low	Low	Low	Low	
Fang 2013	Unclear	Unclear	High	Unclear	Unclear	High	Unclear	Low	
Huipeng 2013	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Cassandra 2014	Low	Unclear	High	Low	High	Low	Low	Low	
Danielsson 2014	Unclear	Low	High	Low	High	Low	Low	Low	
Pfaff 2014	Low	Low	High	Low	Low <sup>1</sup>	Low	Low	High	Baseline difference
Guifeng 2015	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Jinchun 2015	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Schuch 2015	Unclear	Low	High	Low	Low	Low	Low	Low	
Kerling 2015	Unclear	Unclear	High	Unclear	Low	Low	Low	Low	
Belvederi 2015	Low	Low	High	Low	High	Low	Low	High	Post-hoc sample size
Carneiro 2015	Unclear	Low	High	High	Unclear	Low	Low	Low	
Doose 2015	Unclear	Unclear	High	High	High	Low	Low	High	No sample size calc.
Salehi 2016	High	High	High	Low	Unclear	Low	Low	High	Baseline difference
Legrand 2016	Low	High	High	High	High	Low	Unclear	Low	

<sup>1</sup>For the outcome 'lack of remission' this bias domain was high risk of bias

**Table 3.** Heterogeneity of effect estimates for trials assessing the effect of exercise for patients diagnosed with depression explored by comparing sub-groups

Subgroups	Number of Trials (participants)	Random effects meta-analysis SMD (95% CI., p, I <sup>2</sup> )	Subgroup explains heterogeneity P value
<b>Risk of bias</b>			
Less than high risk of bias <sup>1</sup>	4 (530)	-0.11 (-0.41 to 0.18; p = 0.45; I <sup>2</sup> = 62%)	<0.001
High risk of bias	27 (1801)	-0.85 (-1.10 to -0.60; p < 0.001; I <sup>2</sup> = 82%)	
<b>Age</b>			
Old (>59 years)	5 (492)	-0.77 (-1.34 to -0.19; p = 0.009; I <sup>2</sup> = 87%)	0.99
Young (<59 years)	26 (1839)	-0.76 (-1.01 to -0.51; p < 0.001; I <sup>2</sup> = 83%)	
<b>Exercise context</b>			
Group exercise	24 (1729)	-0.79 (-1.06 to -0.52; p < 0.001; I <sup>2</sup> = 85%)	0.72
Individual exercise	7 (602)	-0.68 (-1.17 to -0.20; p = 0.005; I <sup>2</sup> = 79%)	
<b>Duration</b>			
Less than 10 weeks	14 (691)	-0.93 (-1.11 to -0.88; p < 0.001; I <sup>2</sup> = 19%)	0.05
10 weeks or more	17 (1640)	-0.58 (-0.88 to -0.28; p < 0.001; I <sup>2</sup> = 86%)	
<b>Attention control</b>			
Attention control	7 (609)	-0.71 (-1.27 to -0.16; p = 0.01; I <sup>2</sup> = 89%)	0.99
Waitlist	2 (47)	-0.67 (-2.48 to 1.13; p = 0.47; I <sup>2</sup> = 88%)	
<b>Pharmacotherapy</b>			
Add-on	11 (734)	-0.92 (-1.38 to -0.46; p < 0.001; I <sup>2</sup> = 86%)	0.82
No medication	6 (318)	-0.82 (-1.58 to -0.06; p = 0.03; I <sup>2</sup> = 88%)	
<b>Somatic comorbidity</b>			
Somatic co-morbidity	0	N/A	
No co-morbidity	31 (2331)	N/A	
<b>Minor depression</b>			
Incl. minor depression	6 (350)	-0.90 (-1.65 to -0.15; p = 0.02; I <sup>2</sup> = 86%)	0.67
No minor depression	25 (1981)	-0.73 (-0.97 to -0.49; p < 0.001; I <sup>2</sup> = 88%)	
<b>Patient setting</b>			
Inpatients	10 (549)	-0.88 (-1.07 to -0.70; p < 0.001; I <sup>2</sup> = 6%)	0.26
Outpatients	21 (1782)	-0.69 (-0.98 to -0.41; p < 0.001; I <sup>2</sup> = 85%)	
<b>Trial size</b>			
Trials n ≤ 52	15 (479)	-1.30 (-1.74 to -0.86; p < 0.001; I <sup>2</sup> = 77%)	<0.001
Trials n > 52	16 (1852)	-0.40 (-0.60 to -0.19; p < 0.001; I <sup>2</sup> = 76%)	
<b>Increase in exercise capacity</b>			
VO <sub>2</sub> max > 2.6 ml/kg/min	5 (356)	-0.55 (-0.65 to 0.07; p = 0.08; I <sup>2</sup> = 86%)	0.49
VO <sub>2</sub> max ≤ 2.6 ml/kg/min	5 (601)	-0.30 (-0.63 to 0.03; p = 0.07; I <sup>2</sup> = 73%)	

<sup>1</sup>Trials potentially having less bias than trials with high risk of bias.



Table 4. Summary of findings

**Exercise compared to control or treatment as usual for depression**

Patient or population: depression

Setting: In- or out-patients

Intervention: exercise

Comparison: control or treatment as usual

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control or treatment as usual	Risk with exercise				
Severity of depression	-	<b>0.74 SMD lower</b> (0.51 lower to 0.96 lower)	-	2419 (31 RCTs)	⊕○○○ VERY LOW <sup>1</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.
Lack of remission	<b>Study population</b> 646 per 1000	<b>504 per 1000</b> (426 to 594)	<b>RR 0.78</b> (0.68 to 0.90)	1639 (19 RCTs)	⊕○○○ VERY LOW <sup>2</sup>	Remission is, with minor variations, defined as not full-filling the criteria for depression.
Serious adverse events	<b>Study population</b> 0 per 1000	<b>0 per 1000</b> (0 to 0)	<b>RR 2.21</b> (0.24 to 20.21)	335 (3 RCTs)	⊕⊕○○ LOW <sup>3</sup>	
Quality of life	-	<b>0.43 SMD higher</b> (0.04 lower to 0.91 higher)	-	901 (8 RCTs)	⊕○○○ VERY LOW <sup>4</sup>	Quality of life was assessed using a number of different methods. Higher score indicates improved quality of life. Seven of 24 trials reported on this outcome
Depression severity after the intervention	-	<b>0.06 SMD lower</b> (0.25 lower to 0.14 higher)	-	713 (7 RCTs)	⊕⊕○○ LOW <sup>5</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.
Lack of remission after the intervention	<b>Study population</b> 469 per 1000	<b>446 per 1000</b> (385 to 521)	<b>RR 0.95</b> (0.82 to 1.11)	777 (5 RCTs)	⊕⊕○○ LOW <sup>6</sup>	
Depression severity. Restricted to trials with less than high risk of bias.	-	<b>0.11 SMD lower</b> (0.41 lower to 0.18 higher)	-	530 (4 RCTs)	⊕⊕○○ LOW <sup>7</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.

\*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). CI: Confidence interval; SMD: Standardised mean difference; RR: Risk ratio

**GRADE Working Group grades on evidence**

**High quality:** We are very confident that the true effect lies close to that of the estimate of the effect

**Moderate quality:** We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different

**Low quality:** Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect

**low:** We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of the effect

Very

1. Downgraded by 3: risk of bias, inconsistency and publication bias
2. Downgraded by 3: risk of bias, inconsistency and publication bias
3. Downgraded by 2: imprecision and publication bias
4. Downgraded by 3: risk of bias, inconsistency and imprecision

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 5. Downgraded by 2: risk of bias and imprecision
- 6. Downgraded by 2: risk of bias and imprecision
- 7. Downgraded by 2: inconsistency and imprecision

For peer review only

Figure 1. Effect of exercise on depression severity in patients diagnosed with depression

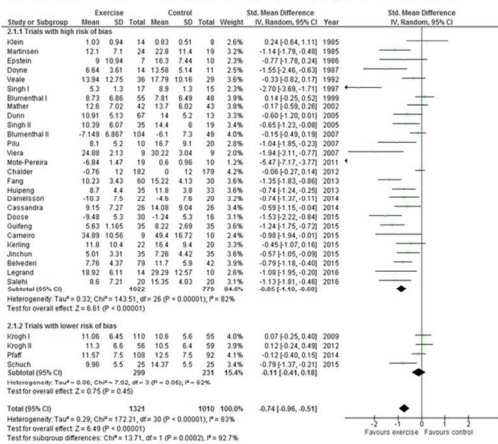


Figure 1

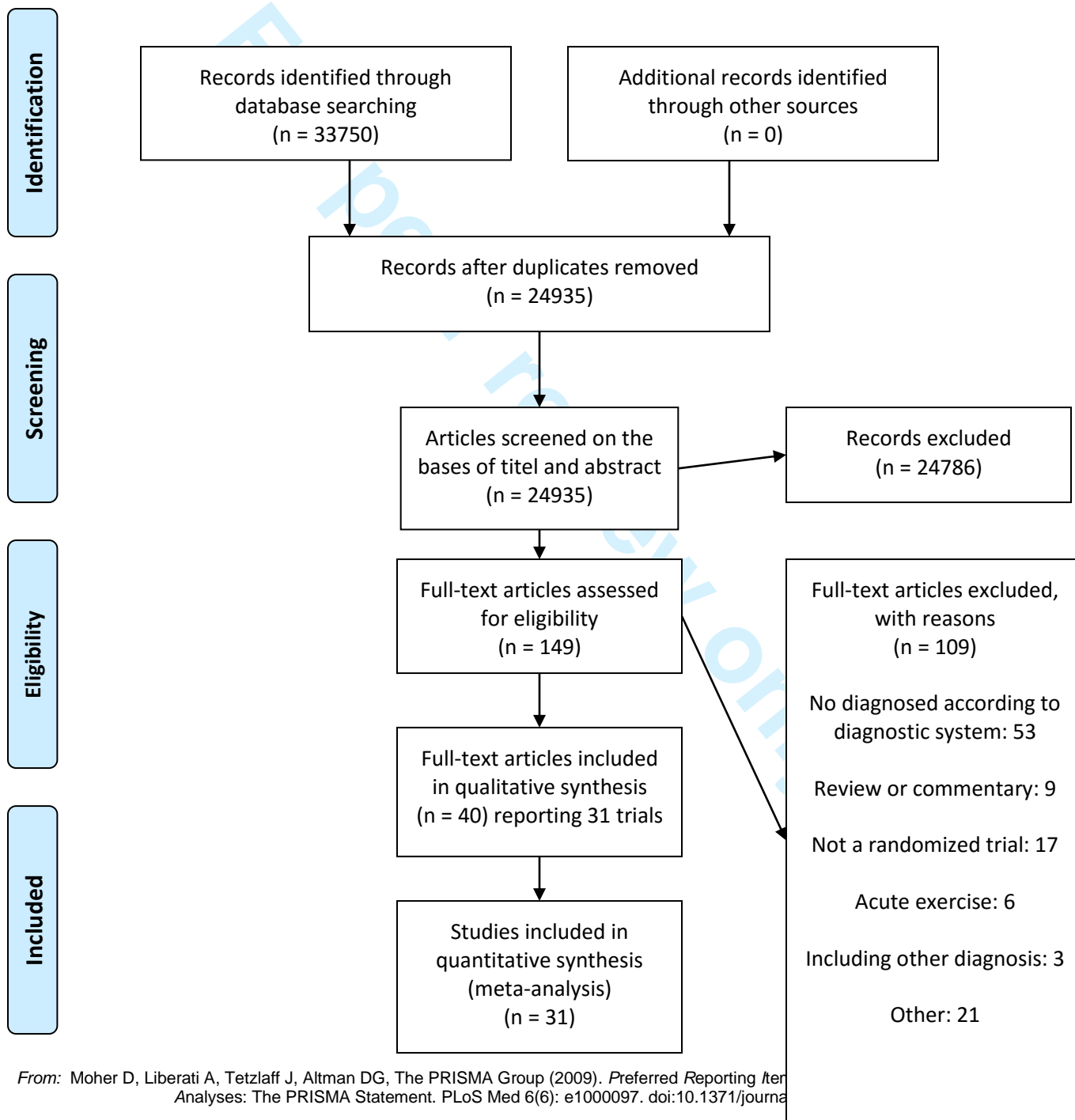
338x190mm (96 x 96 DPI)

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplemental**

Figure S1. Flow diagram for identification of trials assessing the effects of exercise for patients with depression.



BMJ Open: first published as 10.1136/bmjopen-2016-014820 on 18 September 2017. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(6): e1000097. doi:10.1371/journal.pmed.0060161

For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

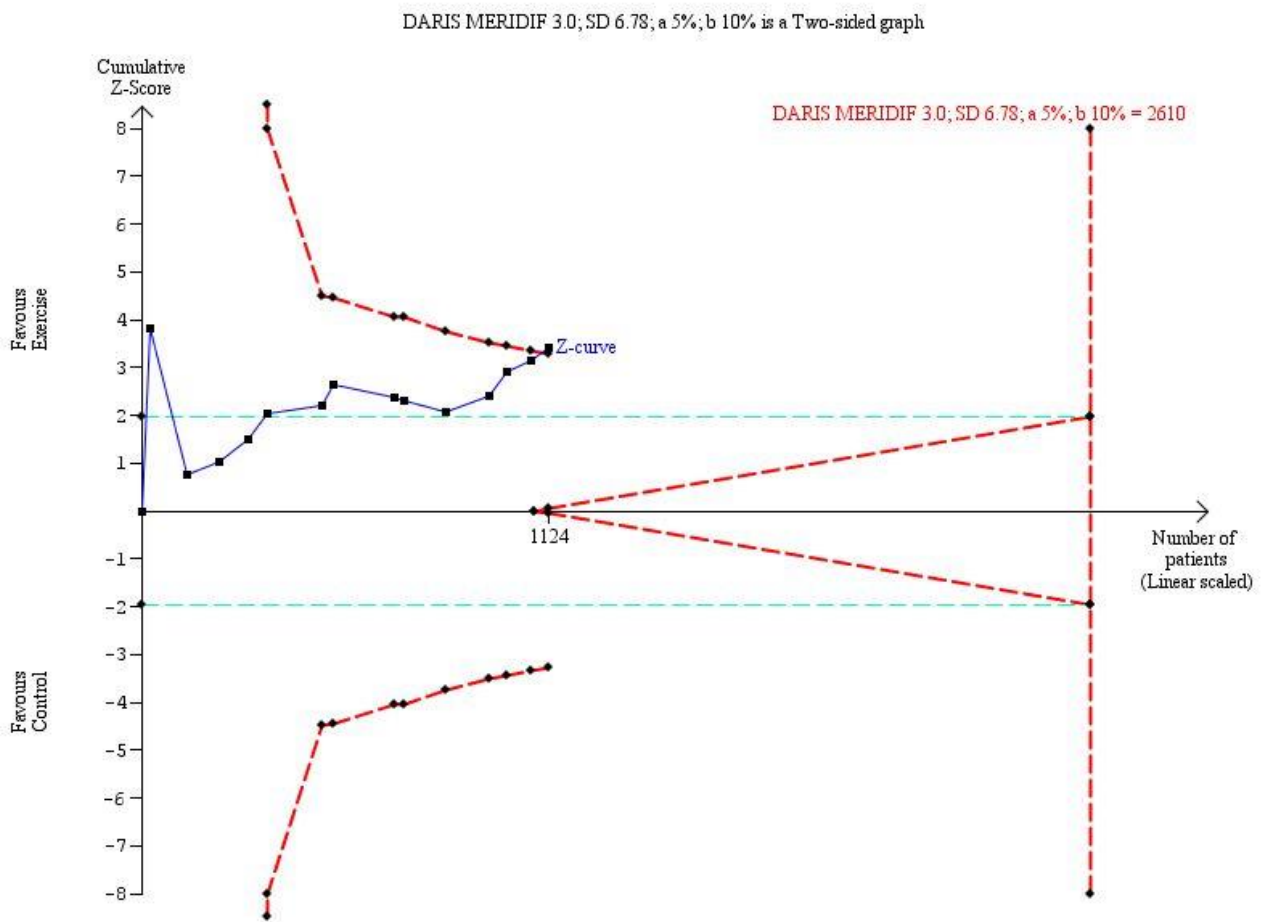
1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S2. Trial Sequential Analysis and required information size for the effect of exercise for depressive symptoms including twelve trials reporting on HAM-D<sub>17</sub>.

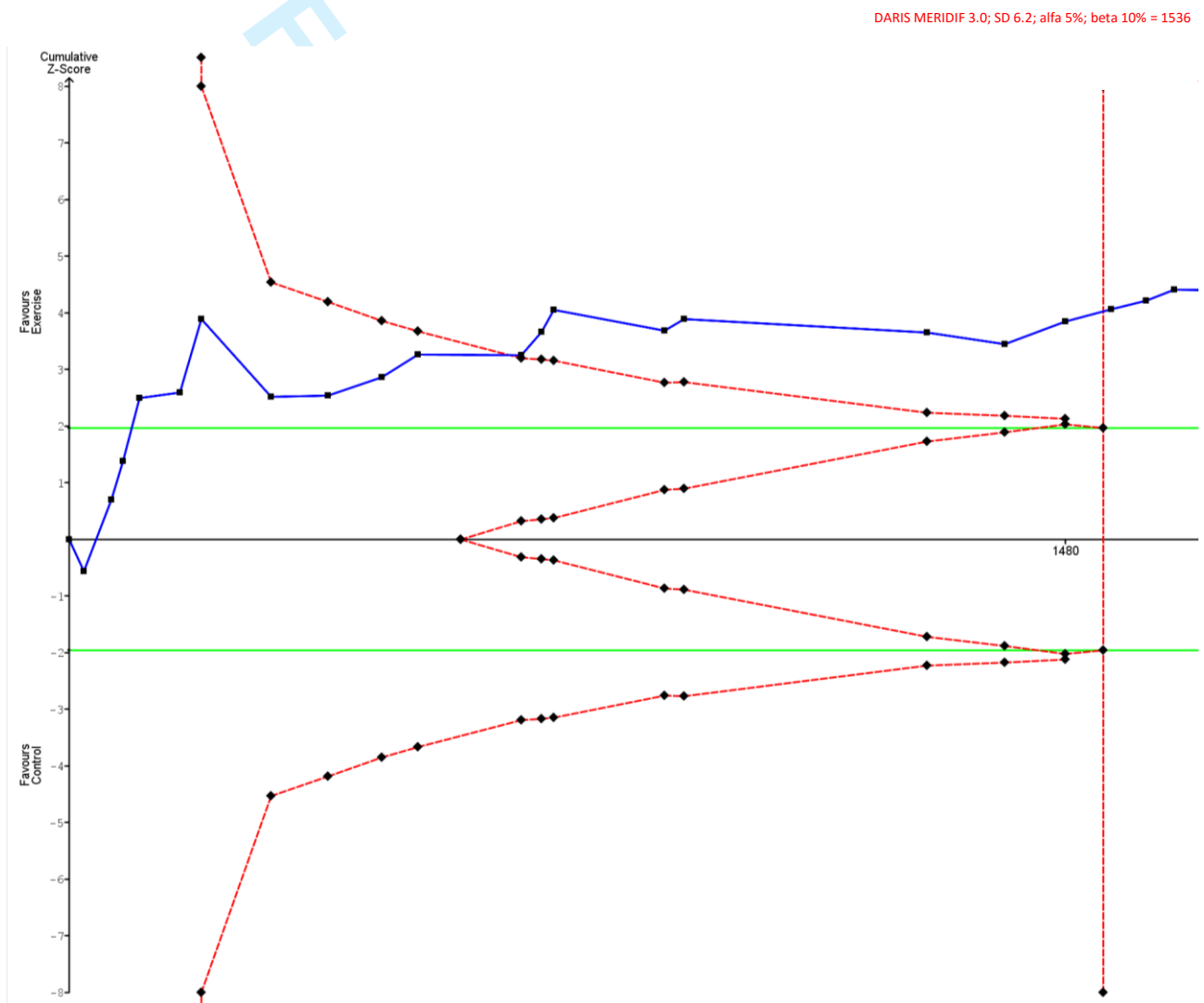


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S3. Trial Sequential Analysis and required information size for the effect of exercise for depressive symptoms including 31 trials 'converted' to a HAM-D<sub>17</sub> scale.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

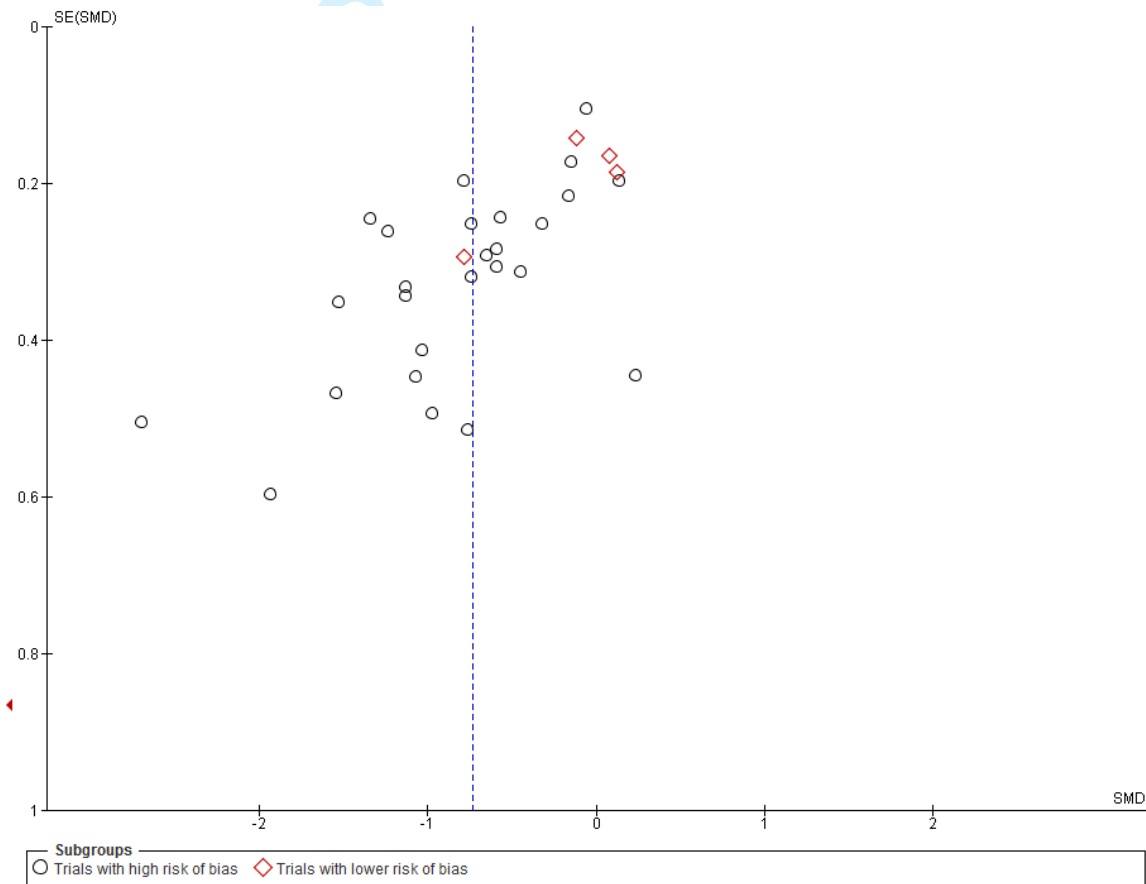
Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Figure**

**Figure S4.**

Funnel plot of 31 trials assessing the antidepressant effect of exercise as a continuous outcome

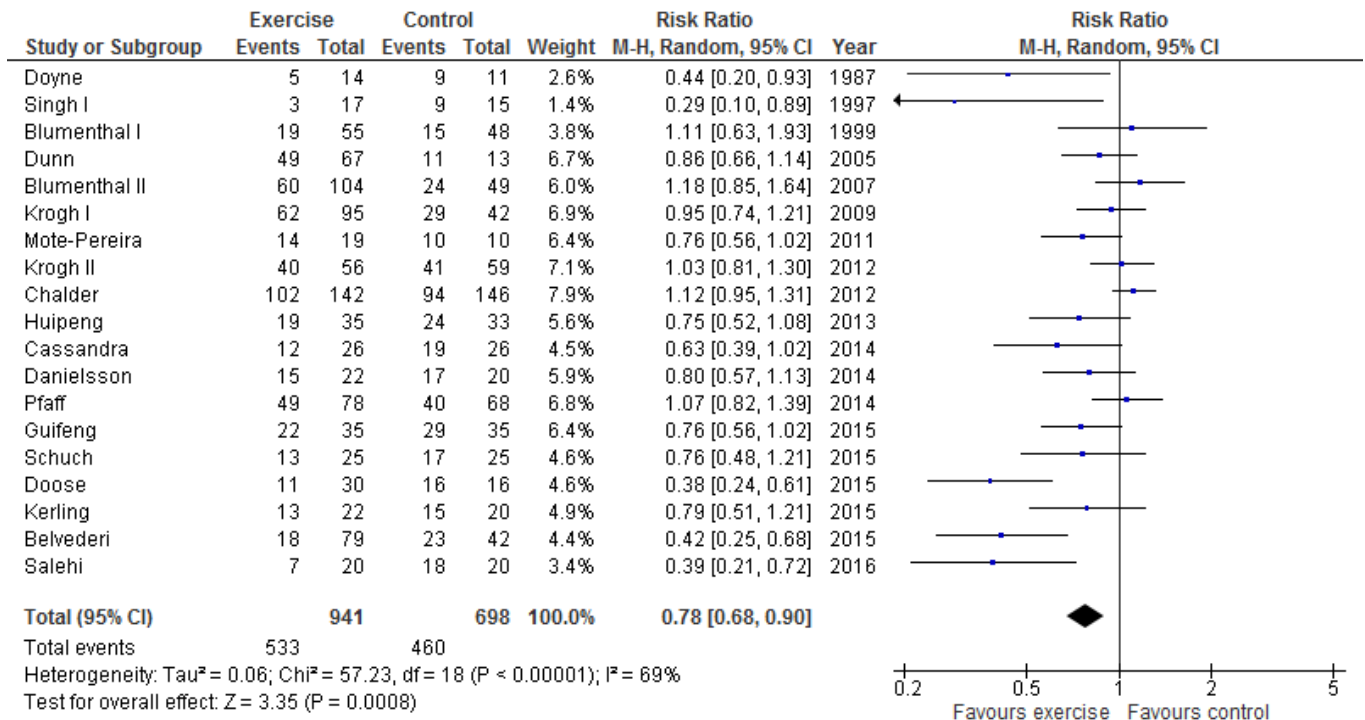


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S5. Effect of exercise on lack of remission for patients diagnosed with depression



View Only

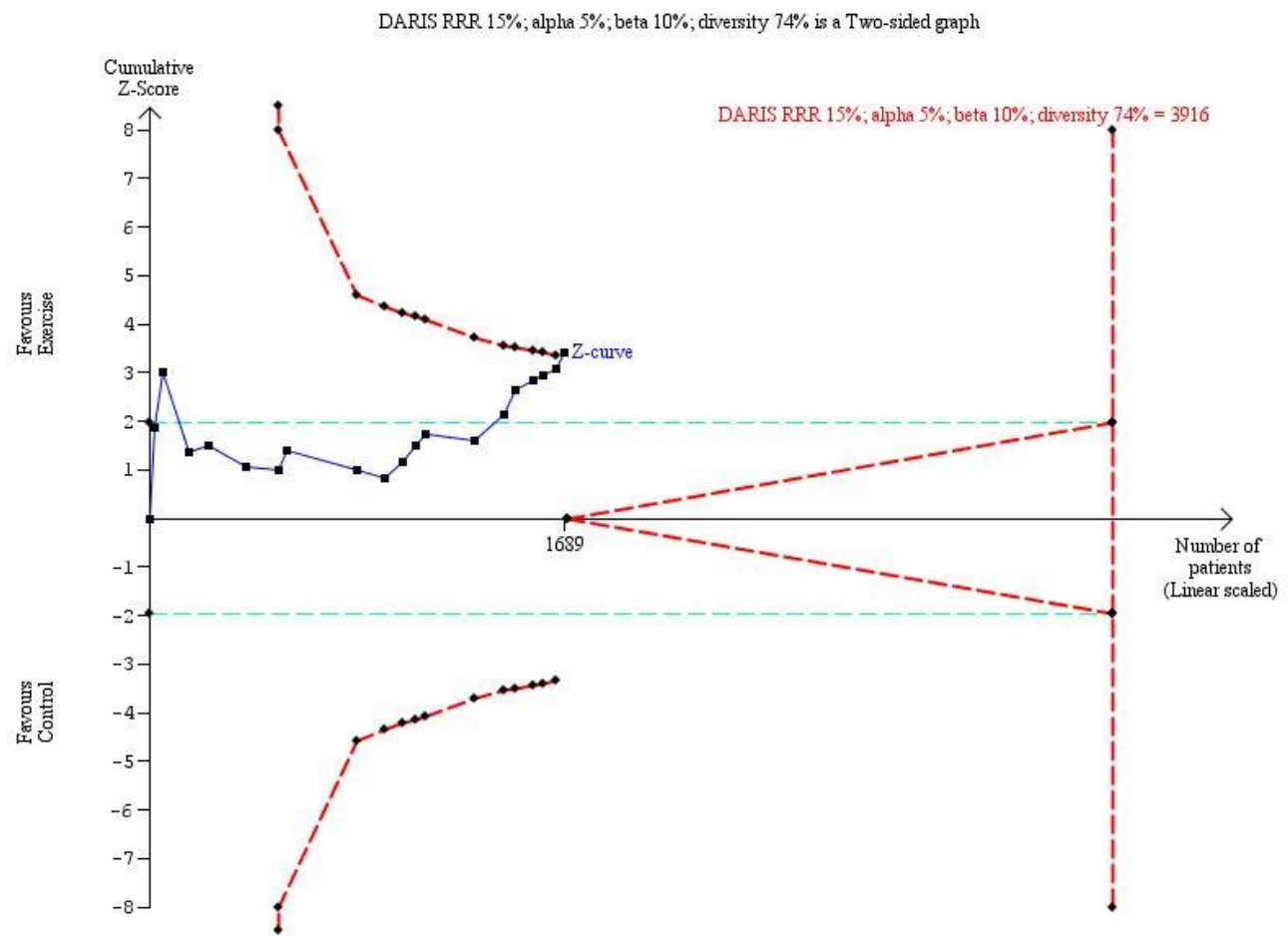


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S6. Trial Sequential Analysis and required information size for the effect of exercise on lack of remission.

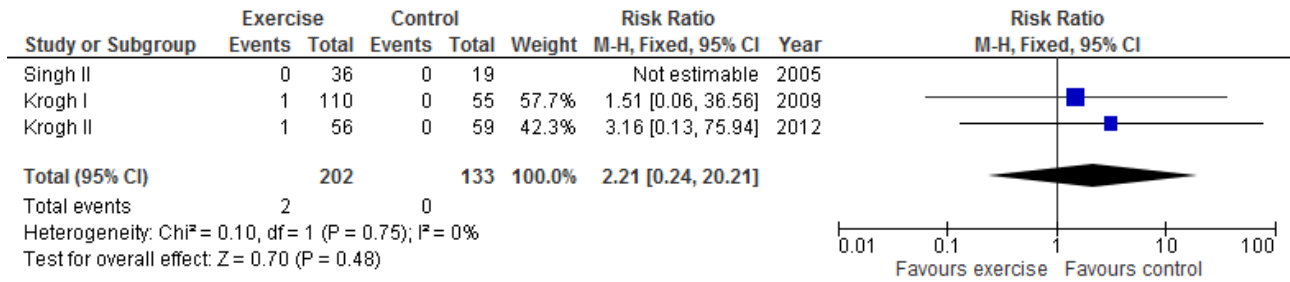


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure S7

Figure S7. Effect of exercise on risk of serious adverse events for patients diagnosed with depression



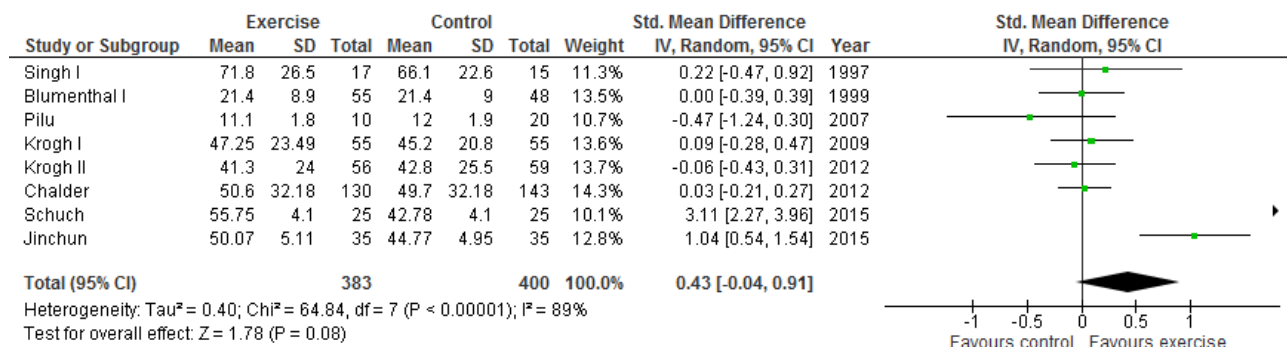
Peer review only

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

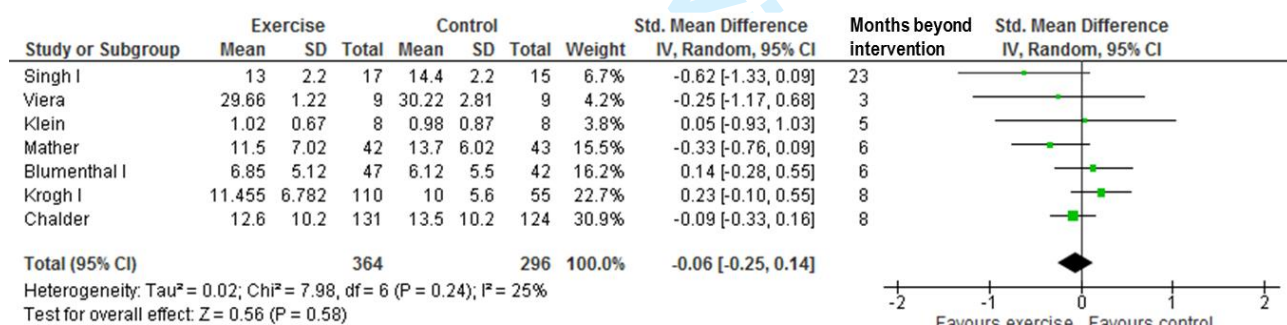
Supplementary Figure S8-S10

Figure S8. The effect of exercise on quality of life in patients diagnosed with depression

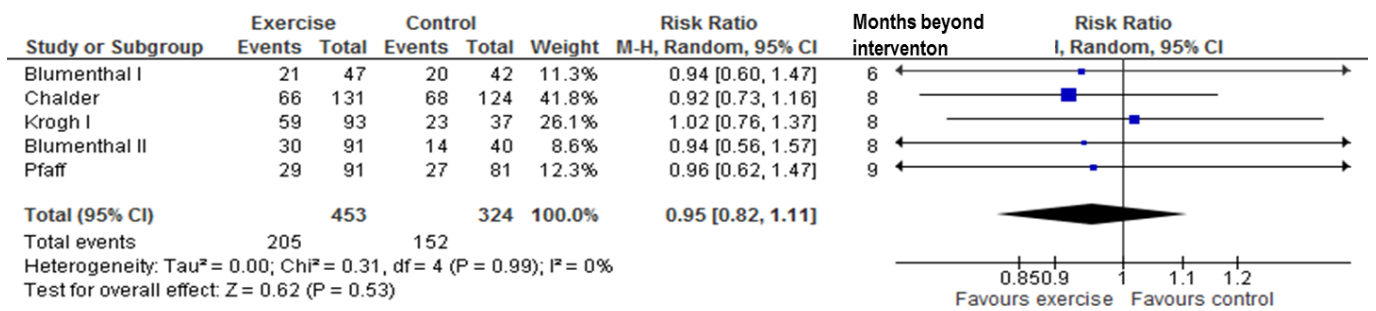


Quality of life was assessed using different scales: Singh I and Chalder used the SF-36, Blumenthal used Life Satisfaction Index, Pilu and Schuch used the WHOQOL, Krogh I and Krogh II used the WHO-Five Well-being Scale, and Jinchun used the GQOLI-74.

Figure S9. The effect of exercise on depression severity after the intervention in patients diagnosed with depression



**Figure S10.** The effect of exercise on risk of lack of remission after the intervention in patients diagnosed with depression



BMJ Open: first published as 10.1136/bmjopen-2016-014820 on 18 September 2017. Downloaded from <http://bmjopen.bmj.com/> on April 18, 2024 by guest. Protected by copyright.

For peer review only

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Table S1**

**Table S1.** Imputation of missing data for trials assessing exercise for patients diagnosed with depression

Outcome	Result from review	Best/worse (1SD)	Best/worse (2SD)	Worse/best (1SD)	Worse/best (2SD)
Depression SMD (95% CI)	-0.74 (-0.96 to -0.51) p < 0.001; I <sup>2</sup> = 83%	-0.85 (-1.10 to -0.60) p < 0.001; I <sup>2</sup> = 87.2%	-0.85 (-1.11 to -0.60) p < 0.001; I <sup>2</sup> = 87.9%	-0.66 (-0.90 to -0.40) p < 0.001; I <sup>2</sup> = 85.4%	-0.61 (-0.84 to -0.38) p < 0.001; I <sup>2</sup> = 85.5%
		<b>Good Outcome</b>	<b>Poor outcome</b>	<b>Good/poor outcome</b>	<b>Poor/good outcome</b>
Lack of remission (95% CL)	RR 0.78 (0.68 to 0.90) p < 0.001; I <sup>2</sup> = 69%	RR 0.75 (0.64 to 0.89) p = 0.0008; I <sup>2</sup> = 73%	RR 0.88 (0.83 to 0.94) p = 0.0002; I <sup>2</sup> = 69%	RR 0.71 (0.61 to 0.81) p < 0.001; I <sup>2</sup> = 68%	RR 0.86 (0.71 to 1.04) p = 0.12; I <sup>2</sup> = 83%
Serious adverse events (95% CL)	RR 2.21 (0.24 to 20.21) p = 0.48; I <sup>2</sup> = 0%	RR 2.19 (0.23 to 20.76) p = 0.50; I <sup>2</sup> = 50%	RR 0.92 (0.37 to 2.30) p = 0.86; I <sup>2</sup> = 60%	RR 0.08 (0.02 to 0.34) p = 0.001; I <sup>2</sup> = 5.4%	RR 19.17 (2.64 to 139.2) p = 0.004; I <sup>2</sup> = 0%

Imputation of missing data for continuous outcome: 'best-worst' - assumed that all participants lost to follow-up in the intervention group had a beneficial outcome (the group mean minus 1 or 2 SD), and all participants lost to follow-up in the placebo group have had a harmful outcome (the group mean plus 1 SD and 2 SD). The reverse 'worst-best-case' scenario is the reverse of the 'best-worst' scenario.

Missing data for the 'remission' outcome was imputed according to the following scenarios: 1) poor outcome analysis: none of the drop-outs/participants lost from both arms experienced the outcome; 2) good outcome analysis: all of the drop-outs/participants lost from both arms experienced the outcome; 3) extreme case analysis favouring the experimental intervention, all of the drop-outs/participants lost from the experimental arm, but none of the drop-outs/participants lost from the control arm experienced the outcome; and 4) extreme case analysis favouring the control: all drop-outs/participants lost from the experimental arm, but none from the control arm experienced the outcome. Missing data for 'serious adverse events' was calculated with the reverse assumptions.

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Table S2**

**Table S2.** Trials characteristics for exploration of heterogeneity in trials assessing the effect of exercise in patients diagnosed with depression

Trial	Lower risk of bias	Age > 60	Group vs. individual	Duration	Attention control waitlist	Exercise as add on to drugs vs. exercise alone	Within-study dose exercise	Increase in VO2max <sup>1</sup>	Somatic disease vs. only MD	Trial Includes minor depression
Klein 1985	No	Young	Individual	12 weeks	Other	Exercise alone	No	No	No	Yes
Martinsen 1985	No	Young	Group	9 weeks	Attention control	Unclear	No	11 <sup>a</sup>	No	No
Epstein 1986	No	Young	Group	8 weeks	Waitlist	Unclear	No	No	No	Yes
Doyne 1987	No	Young	Individual	8 weeks	Waitlist	Exercise alone	No	No	No	Yes
Veale 1992	No	Young	Group	12 weeks	Other	Unclear	No	No	No	No
Singh 1997	No	Old	Group	10 weeks	Attention control	Exercise alone	No	N/A	No	Yes
Blumenthal 1999	No	Young	Group	16 weeks	Other	Add on	No	2.3	No	No
Mather 2002	No	Old	Group	10 weeks	Attention control	Add on	No	No	No	No
Dunn 2005	No	Young	Individual	12 weeks	Attention control	Exercise alone	Yes	No	No	No
Singh 2005	No	Old	Group	8 weeks	Other	Exercise alone	Yes	N/A	No	Yes
Pilu 2007	No	Young	Group	24 weeks	Other	Add on	No	No	No	No
Viera 2007	No	Young	Group	12 weeks	Other	Add on	No	No	No	No
Blumenthal 2007	No	Young	Group	16 weeks	Other	Add on	No	2.0 <sup>a</sup>	No	No
Krogh 2009	Yes	Young	Group	16 weeks	Attention control	No	No	2.9	No	No
Mota-Pereira 2011	No	Young	Group	12 weeks	Other	Add on	No	No	No	No
Krogh 2012	Yes	Young	Group	12 weeks	Attention control	Exercise alone	No	3.4	No	No
Chalder 2012	No	Young	Individual	32 weeks	Other	No	No	No	No	No
Fang 2013	No	Young	Group	6 weeks	Attention control	No	Yes	No	No	No
Huipeng 2013	No	Young	Group	6 weeks	Other	No	No	No	No	No
Cassandra 2014	No	Young	Group	3 weeks	Other	Add on	No	No	No	No
Danielsson 2014	No	Young	Group	10 weeks	Other	Add on	No	2.4	No	No
Pfaff 2014	Yes	Old	Group	12 weeks	Other	No	No	1.5	No	Yes
Guifeng 2015	No	Young	Group	8 weeks	Other	No	No	No	No	No
Jinchun 2015	No	Young	Group	8 weeks	Other	No	No	No	No	No
Schuch 2015	Yes	Young	Individual	2 weeks	Other	No	No	No	No	No
Kerling 2015	No	Young	Group	6 weeks	Other	No	No	2.8	No	No

1	Belvederi	No	Old	Group	24 weeks	Other	Add on	Yes	0.3 <sup>a</sup>	No	No
2	2015										
3	Carneiro	No	Young	Group	16 weeks	Other	Add on	No	No	No	No
4	2015										
5	Doose	No	Young	Group	8 weeks	Other	No	No	3.2	No	No
6	2015										
7	Legrand	No	Young	Individual	10 days	Other	No	No	No	No	No
8	2016										
9	Salehi	No	Young	Individual	4 weeks	Other	Add on	No	No	No	No
10	2016										

<sup>1</sup>Increase in VO2max is based on increase in intervention group, if <sup>a</sup> then value is based on an estimate from text or figures.

For peer review only

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplementary Table

**Table S3.** Heterogeneity of effect estimates for trials assessing the effect of exercise for patients diagnosed with depression on lack of remission.

Subgroups	Number of Trials (participants)	Random effects meta-analysis RR (95% CI., p, I <sup>2</sup> )	Subgroup explains heterogeneity P value
<b>Risk of bias</b>			
Less than high risk of bias <sup>1,2</sup>	2 (165)	0.95 (0.74 to 1.23; p = 0.70; I <sup>2</sup> = 20%)	0.18
High risk of bias	17 (1474)	0.77 (0.64 to 0.92; p = 0.003; I <sup>2</sup> = 75%)	
<b>Age</b>			
Old (>59 years)	3 (299)	0.61 (0.21 to 1.02; p = 0.37; I <sup>2</sup> = 91%)	0.62
Young (<59 years)	16 (1340)	0.81 (0.70 to 0.93; p = 0.003; I <sup>2</sup> = 64%)	
<b>Exercise context</b>			
Group exercise	14 (1156)	0.80 (0.66 to 0.96; p = 0.02; I <sup>2</sup> = 72%)	0.69
Individual exercise	5 (483)	0.74 (0.52 to 1.04; p = 0.08; I <sup>2</sup> = 77%)	
<b>Duration</b>			
Less than 10 weeks	8 (393)	0.63 (0.51 to 0.77; p < 0.001; I <sup>2</sup> = 40%)	0.004
10 weeks or more	11 (1246)	0.93 (0.78 to 1.10; p = 0.39; I <sup>2</sup> = 69%)	
<b>Attention control</b>			
Attention control	4 (364)	0.91 (0.73 to 1.12; p = 0.38; I <sup>2</sup> = 42%)	0.07
Waitlist	1 (25)	0.44 (0.21 to 0.93; p = 0.03; I <sup>2</sup> = 0%)	
<b>Pharmacotherapy</b>			
Add-on	7 (540)	0.72 (0.54 to 0.96; p = 0.03; I <sup>2</sup> = 69%)	0.62
No medication	4 (252)	0.75 (0.52 to 1.09; p = 0.13; I <sup>2</sup> = 66%)	
<b>Somatic comorbidity</b>			
Somatic co-morbidity	0	N/A	
No co-morbidity	19 (1639)	N/A	
<b>Minor depression</b>			
Incl. minor depression	3 (203)	0.63 (0.21 to 1.89; p = 0.41; I <sup>2</sup> = 87%)	0.69
No minor depression	16 (1436)	0.79 (0.68 to 0.92; p = 0.002; I <sup>2</sup> = 69%)	
<b>Patient setting</b>			
Inpatients	6 (322)	0.71 (0.60 to 0.84; p < 0.001; I <sup>2</sup> = 0%)	0.21
Outpatients	13 (1317)	0.84 (0.69 to 1.01; p = 0.07; I <sup>2</sup> = 77%)	
<b>Trial size</b>			
Trials n ≤ 52	9 (358)	0.62 (0.50 to 0.76; p < 0.001; I <sup>2</sup> = 45%)	0.002
Trials n > 52	10 (1281)	0.95 (0.80 to 1.12; p = 0.52; I <sup>2</sup> = 68%)	

<sup>1</sup>Trials potentially having less bias than trials with high risk of bias.



1  
2  
3  
4 Article:

5  
6 Exercise for patients with major depression: A systematic review with meta-analysis and trial  
7 sequential-analysis  
8

9  
10 **Supplementary material (S1)**

11 An example of bibliographical search for PubMed

12 #1 Depression [MeSH]

13 #2 Depressive disorder [MeSH]

14 #3 Exercise [Text Word]

15 #4 Aerobic [Text Word]

16 #5 Non-aerobic [Text Word]

17 #6 Physical activity [Text Word]

18 #7 Physical fitness [Text Word]

19 #8 Walking [MeSH]

20 #9 Jogging [MeSH]

21 #10 Running [MeSH]

22 #11 Bicycling [MeSH]

23 #12 Swimming [MeSH]

24 #13 Strength [Text Word]

25 #14 Resistance [Text Word]

26 #15 #1 OR #2

27 #16 #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14

28 #17 #15 AND #16  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	1
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ for each meta-analysis).	9



# PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	11
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig. 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Fig 3-fig8
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Table 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Table 3
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	21
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	22
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	26

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org).

Page 2 of 2

# BMJ Open

## Exercise for patients with major depression: a systematic review with meta-analysis and Trial Sequential Analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2016-014820.R2
Article Type:	Research
Date Submitted by the Author:	26-Jun-2017
Complete List of Authors:	Krogh, Jesper; Kobenhavns Universitet Sundhedsvidenskabelige Fakultet, Mental Health Centre Copenhagen Hjorthøj, Carsten; Copenhagen University Hospital, Mental Health Center Copenhagen Speyer, Helene; University of Copenhagen Gluud, Christian; Copenhagen University Hospital Rigshospitalet, The Copenhagen Trial Unit, Centre for Clinical Intervention Research Nordentoft, Merete; University of Copenhagen, Mental Health Centre Copenhagen
<b>Primary Subject Heading</b>:	Mental health
Secondary Subject Heading:	Sports and exercise medicine, Evidence based practice
Keywords:	Exercise, Systematic review, Meta-analysis, Evidence based medicine, Randomised clinical trials, Depression & mood disorders < PSYCHIATRY

SCHOLARONE™  
Manuscripts

Only

1  
2  
3  
4 **Exercise for patients with major depression: a systematic review with meta-**  
5 **analysis and Trial Sequential Analysis**  
6

7  
8 Jesper Krogh<sup>1</sup>, Carsten Hjorthøj<sup>1</sup>, Helene Speyer<sup>1</sup>, Christian Gluud<sup>1</sup>, Merete  
9 Nordentoft<sup>1</sup>  
10

11  
12  
13 <sup>1</sup> Mental Health Centre Copenhagen. Faculty of Health Sciences. University of Copenhagen, Copenhagen,  
14 Denmark.

15  
16 <sup>2</sup> Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University  
17 Hospital, Copenhagen, Denmark.  
18  
19

20  
21  
22  
23 **Corresponding author**

24 Jesper Krogh, [Jesper.krogh@dadlnet.dk](mailto:Jesper.krogh@dadlnet.dk)

25  
26 Mental Health Centre Copenhagen, Kildegårdsvej 28, Building 15. DK-2900, Hellerup, Denmark  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Abstract

### Objectives

To assess the benefits and harms of exercise in patients with depression.

### Design

Systematic review

### Data sources

Bibliographical databases were searched until the 20<sup>th</sup> of June, 2017.

### Eligibility criteria and outcomes

Eligible trials were randomised clinical trials assessing the effect of exercise in participants diagnosed with depression. Primary outcomes were depression severity, lack of remission, and serious adverse events (e.g. suicide) assessed at the end of the intervention. Secondary outcomes were quality of life and adverse events such as injuries, as well as assessment of depression severity and lack of remission during follow-up after the intervention.

### Results

Thirty-five trials enrolling 2498 participants were included. The effect of exercise versus control on depression severity was -0.66 standardised mean difference (SMD) (95% CI -0.86 to -0.46;  $P < 0.001$ ; GRADE: very low quality). Restricting this analysis to the four trials that seemed less affected of bias, the effect vanished into -0.11 SMD (-0.41 to 0.18;  $P = 0.45$ ; GRADE: low quality). Exercise decreased the relative risk of no remission to 0.78 (0.68 to 0.90;  $P < 0.001$ ; GRADE: very low quality). Restricting this analysis to the two trials that seemed less affected of bias, the effect vanished into 0.95 (0.74 to 1.23;  $P = 0.78$ ). Trial Sequential Analysis excluded random error when all trials were analysed, but not if focusing on trials less affected of bias. Sub-group analyses found that trial size and intervention duration were inversely

1  
2  
3  
4 associated with effect size for both depression severity and lack of remission. There was no significant  
5  
6 effect of exercise on secondary outcomes.  
7  
8

### 9 **Conclusions**

10  
11 Trials with less risk of bias suggested no antidepressant effects of exercise and there were no significant  
12  
13 effects of exercise on quality of life, depression severity, or lack of remission during follow-up. Data for  
14  
15 serious adverse events and adverse events was scarce not allowing conclusions for these outcomes.  
16  
17

### 18 **Systematic review registration**

19  
20  
21 The protocol was published in the journal Systematic Reviews: 2015; 4:40.  
22  
23

24 **DOI:** 10.1186/s13643-015-0030-6.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Article Summary

### Strengths and limitations of this study

- The protocol for this review has previously been published
- Using meta-regression analysis, Trial Sequential Analysis and the GRADE system the conclusions from this review is based on a firm and transparent platform
- Based on an extensive literature search, this review included 35 trials allocating almost 2500 participants diagnosed with depression to exercise or control interventions than could be analysed
- The effect estimates are largely based on trials at high risk of bias
- Effect estimates from included trials had considerable heterogeneity



## Introduction

Depression is a common disorder affecting up to 17% of the population during their lifetime.<sup>1;2</sup> Based on data from the World Health Organisation, depression is ranked as the second largest health-care problem globally, in terms of years lived with disability.<sup>3</sup> Depending on its severity, depression is often treated using psychotherapy, antidepressants, or a combination of both. However, the clinical benefits of antidepressants<sup>4-6</sup> and psychotherapy<sup>7-9</sup> has been challenged. Both treatments are costly in terms of time and money and may also have adverse effects. Compliance with antidepressant treatment is poor; the dropout rate in clinical trials is reported to be between 12% and 40% within the initial 6 to 8 weeks of treatment.<sup>4;10</sup>

The weakness of evidence for the beneficial effect of current interventions, along with problems related to low compliance and harms, has resulted in an interest in using alternative interventions. The use of exercise as an intervention has attracted considerable attention, and various forms of exercise varying in intensity have been assessed in a number of randomised clinical trials to test their effectiveness as a treatment for patients with depression. In 2011, we published a meta-analysis of randomised clinical trials examining the effect of exercise on depressive symptoms in patients with clinical depression.<sup>11</sup> The results suggested that referring patients with clinical depression to exercise programs was associated with a small to moderate effect on depressive symptoms. However, restricting the analysis to three trials at low risk of bias, the effect estimate was non-significant. Since 2011, other reviews have been published on the effect of exercise on depressive symptoms,<sup>12</sup> in older people,<sup>13</sup> and in patients with chronic illnesses.<sup>14</sup> However, none of these reviews addressed the specific population of adults diagnosed with major depression according to valid diagnostic criteria, such as the International Classification of Diseases<sup>15</sup> or the Diagnostic and Statistical Manual of Mental Disorders.<sup>16</sup> The reviews contained a number of trials that included volunteers who were defined as being depressed on the basis of psychometric testing (for example, Beck

1  
2  
3  
4 Depression Inventory<sup>17</sup>), as opposed to individuals with a clinical diagnosis of major depression.

5  
6 Furthermore, several randomised clinical trials investigating the effect of exercise in clinically depressed  
7  
8 individuals have been published since our 2011 review.<sup>11</sup>  
9

10  
11  
12  
13  
14 The objectives of the present systematic review are to investigate the beneficial and harmful effects of  
15  
16 exercise, in terms of severity of depression, lack of remission, quality of life, and suicide versus controls  
17  
18 with or without co-interventions in adults with a clinical diagnosis of major depression. The current  
19  
20 systematic review differs from our previous review in a number of aspects.<sup>11</sup> We only considered trials  
21  
22 including participants diagnosed with depression according to a validated diagnostic system. We also  
23  
24 included trials including participants with somatic co-morbidity, e.g. cancer or diabetes. The harmful effects  
25  
26 of exercise interventions are also addressed, the intervention effects being assessed according to the  
27  
28 grading of recommendations assessment, development, and evaluation (GRADE) framework, and  
29  
30 bibliographical searches have been extended to include a Chinese and a South-American database until  
31  
32 2016.  
33  
34  
35  
36  
37  
38

### 39 **Methods/design**

40  
41  
42 The protocol for this review has previously been published.<sup>18</sup>  
43  
44  
45  
46

### 47 **Search strategy**

48  
49  
50 The following bibliographical databases was searched: CENTRAL, MEDLINE, EMBASE, Science Citation Index  
51  
52 (Web of Science), LILACS, and Wanfang using medical subject headings (MeSH or similar) when possible or  
53  
54 text word terms: depression, depressive disorder and exercise, aerobic, non-aerobic, physical activity,  
55  
56 physical fitness, walking, jogging, running, bicycling, swimming, strength, or resistance. Please see  
57  
58  
59  
60

1  
2  
3  
4 supplementary material (S1) for an example of a bibliographical search. The main search was conducted in  
5  
6 August 2015, and the latest search was conducted on 20<sup>th</sup> of June, 2017.  
7  
8  
9

## 10 11 **Trial selection**

12  
13  
14 One investigator (JK) examined titles and abstracts to remove obviously irrelevant reports. Two  
15  
16 investigators (JK + HS) examined full text reports and abstracts determining compliance with inclusion  
17  
18 criteria. A trial was considered eligible if it was a randomised clinical trials including participants diagnosed  
19  
20 as having major depression according to a valid and recognised diagnostic system (that is, Research  
21  
22 Diagnostic Criteria (RDC),<sup>19</sup> International Classification of Diseases (ICD)<sup>15</sup> or Diagnostic and Statistical  
23  
24 Manual of Mental disorders (DSM)<sup>16</sup>) and included participants aged >17 years. Abstracts and full text  
25  
26 reports were included.  
27  
28  
29

30  
31  
32  
33 Trials were excluded if they measured depression immediately after a single bout of exercise, compared  
34  
35 one form of exercise versus another, or compared different exercise intensities without including a control  
36  
37 group. The trials had to allocate participants to an exercise intervention versus a control group (that is,  
38  
39 exercise versus a control group receiving no intervention or treatment as usual or an attention control  
40  
41 using light exercise) or using exercise as an add-on-treatment (that is, exercise plus usual treatment in the  
42  
43 experimental group versus usual treatment alone in the control group). Exercise intervention was defined  
44  
45 as a systematic physical intervention with the intention to increase muscle strength and/or cardiovascular  
46  
47 fitness, e.g. running, swimming or weight lifting. In case of attention control, it should specifically be  
48  
49 mentioned by the authors of the trial report that the intervention was intended as a control intervention.  
50  
51  
52

## 53 54 55 **Outcomes**

1  
2  
3  
4 The primary outcomes were 1) depressive symptoms measured on a continuous scale assessed at the end  
5  
6 of the intervention; 2) lack of remission, that is, a binary outcome of the proportion of participants in each  
7  
8 intervention group of the trial who did not obtain remission at the end of the intervention according to the  
9  
10 authors' own definition; and 3) serious adverse events defined according to ICH-GCP as any untoward  
11  
12 medical occurrence that was life threatening, resulted in death or persistent or significant disability (ICH-  
13  
14 GCP 1997).<sup>20</sup> Serious adverse events accordingly include suicide attempts as well as suicides. The secondary  
15  
16 outcomes were quality of life, non-serious adverse events (e.g. muscle injuries) as well as depressive  
17  
18 symptoms and lack of remission assessed after the intervention.  
19  
20  
21  
22  
23  
24

### 25 **Data extraction**

26  
27 Two authors (JK, HS) independently extracted data using a pre-piloted structured form. Any discrepancies  
28  
29 in the data extraction or inclusion/exclusion of trials was resolved by referring to the original papers. CG or  
30  
31 MN assisted as adjudicator in cases of disagreements. Data extraction included, in addition to outcomes,  
32  
33 information regarding country of origin, number of randomised participants, number of participants  
34  
35 included in efficacy analysis, mean age of participants, diagnostic system, baseline assessment of  
36  
37 depression severity, type of intervention, frequency of intervention, and duration of intervention.  
38  
39 Continuous outcomes were preferred in the following order: post-intervention scores with corresponding  
40  
41 standard deviations (SD), mean change from baseline with SD, mean difference between groups post-  
42  
43 intervention and reported outcomes were preferred to figures. JK and CH independently performed the  
44  
45 assessment of bias domains. The authors JK, CG, and MN have previously published trial reports assessing  
46  
47 the effect of exercise in participants with depression,<sup>21;22</sup> and to reduce the risk of academic bias two  
48  
49 additional authors were included in the current systematic review (CH, HS).  
50  
51  
52  
53  
54  
55  
56  
57

### 58 **Risk of bias assessment**

1  
2  
3  
4 Definitions in the assessment of bias risk of a trial was conducted according to the Cochrane Handbook for  
5  
6 Systematic Reviews of Interventions<sup>23</sup> of the following domains: allocation sequence generation, allocation  
7  
8 concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome  
9  
10 data, selective outcome reporting, for-profit bias, and other bias. Trials assessed as having 'low risk of bias'  
11  
12 in all of the above specified domains were considered 'trials at low risk of bias'. Trials assessed as having  
13  
14 'uncertain risk of bias' or 'high risk of bias' in one or more of the above specified domains were considered  
15  
16 trials at 'high risk of bias'. In line with our previous systematic review<sup>11</sup> and the latest Cochrane review on  
17  
18 exercise for depression,<sup>24</sup> trials at low risk of bias in the allocation concealment domain, blinded outcome  
19  
20 assessment domain, and the incomplete outcome data domain were characterised as 'trials potentially  
21  
22 having less risk of bias than other trials at high risk of bias'. Trials assessing the effect of behavioural  
23  
24 interventions are rarely able to mask the allocation, and participants and health care providers are  
25  
26 therefore not blinded. Therefore, we will also report the number of trials at low risk of bias in the remaining  
27  
28 domains.  
29  
30  
31  
32  
33  
34  
35

### 36 **Data synthesis and analysis**

37  
38 In order to be able to include all of the trials in our meta-analysis, estimates of standardised mean  
39  
40 difference (SMD) for each individual trial was carried out. SMD is the mean difference in depression score  
41  
42 between the exercise and control groups divided by the pooled standard deviation at follow-up. The result  
43  
44 is a unit free effect size. By convention, SMD effect sizes of 0.2, 0.5 and 0.8 are considered small, medium  
45  
46 and large intervention effects.<sup>23</sup> For dichotomous variables, we calculated the risk ratio (RR) with a 95%  
47  
48 confidence interval. It was expected that some trials would have several intervention groups. Data from the  
49  
50 experimental groups was pooled and compared with the data from the control group. In case of  
51  
52 discrepancies between the random-effects model analysis and the fixed-effect model analysis, both results  
53  
54 are reported; otherwise, only results from the random-effects analysis is reported. The degree of  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 heterogeneity was quantified using the I-squared statistic,<sup>25</sup> which can be interpreted as the percentage of  
5  
6 variation observed between the trials attributable to between-trial differences, rather than sampling error  
7  
8 (chance). Heterogeneity was explored by analyses of sub-groups (see below).  
9

10  
11  
12  
13  
14 For the primary outcomes, Trial Sequential Analysis was performed.<sup>26,27</sup> In order to calculate the required  
15  
16 information size and the cumulative Z-curve's eventual breach of relevant trial sequential monitoring  
17  
18 boundaries, the required information size for the primary continuous outcome was based on type I error of  
19  
20 5%, a beta of 10%, the standard error of the meta-analysis, and a minimal difference of three points on the  
21  
22 HAM-D<sub>17</sub>.<sup>18</sup> Post-hoc we calculated the required information size including all trials. This was done by  
23  
24 converting effect estimates from trials reporting other outcome scales into the HAM-D<sub>17</sub> scale as described  
25  
26 by Thorlund et al.<sup>28</sup> In order to calculate the required information size and the cumulative Z-curve's  
27  
28 eventual breach of relevant trial sequential monitoring boundaries, the required information size for lack of  
29  
30 remission was based on type I error of 5%, a beta of 10%, the proportion of participants in the control  
31  
32 group with the outcome, and a relative risk reduction of 15% and 30%.  
33  
34  
35  
36  
37  
38

39 Bayes factors were calculated for all primary outcomes.<sup>29</sup> Low P-values suggest that we can reject the null-  
40  
41 hypothesis. But even a low P-value from a meta-analysis can be misleading if there is also a low probability  
42  
43 that data are compatible with the anticipated intervention effect. In other words, the probability that the  
44  
45 actual measured difference in effect of the compared interventions resulted from an a priori anticipated  
46  
47 'true' difference needs to be considered. For this purpose, it is helpful to calculate the Bayes factor, which  
48  
49 is the ratio of the P-value probabilities of the meta-analysis result divided by the probability of the  
50  
51 anticipated effect, or 'true' effect.<sup>29</sup> As suggested by Jakobsen et al.,<sup>29</sup> a Bayes factor lower than 0.1  
52  
53 together with a low P-value suggest, if bias can be ruled out, that the observed result is compatible with the  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 a priori expected effect. If the Bayes factor is higher than 0.1 the result is not compatible with the a priori  
5  
6 expected effect and the effect may be lower.  
7  
8  
9

10  
11 To assess the potential impact of missing data (incomplete outcome data bias) we did sensitivity analysis of  
12  
13 missing data using the following strategy: a 'best-worst' case scenario was assessed, assuming that all  
14  
15 participants lost to follow-up in the intervention group had a beneficial outcome (the group mean minus 1  
16  
17 standard deviation (SD)), and all those with missing outcomes in the control group have had a harmful  
18  
19 outcome (the group mean plus 1 SD and 2 SD). In addition, the reverse 'worst-best-case' scenario analysis  
20  
21 was also performed.<sup>29</sup> Missing data for the 'lack of remission' outcome was imputed in sensitivity analysis  
22  
23 according to the following scenarios:<sup>30</sup> 1) poor outcome analysis: assuming that all of the drop-  
24  
25 outs/participants lost from both the experimental and the control arms experienced the outcome, including  
26  
27 all randomised participants in the denominator; 2) good outcome analysis: assuming that none of the drop-  
28  
29 outs/participants lost from the experimental and the control arms experienced the outcome, including all  
30  
31 randomised participants in the denominator; 3) extreme case analysis favouring the experimental  
32  
33 intervention ('best-worse' case scenario): none of the drop-outs/participants lost from the experimental  
34  
35 arm, but all of the drop-outs/participants lost from the control arm experienced the outcome, including all  
36  
37 randomised participants in the denominator; and 4) extreme case analysis favouring the control ('worst-  
38  
39 best' case scenario): all of the drop-outs/participants lost from the experimental arm, but none from the  
40  
41 control arm experienced the outcome, including all randomised participants in the denominator.  
42  
43  
44  
45  
46  
47  
48  
49

### 50 Subgroup analyses

51  
52 In subgroup analyses, the possible effects of variables on intervention effects on outcomes and  
53  
54 heterogeneity were compared. Trials potentially having less risk of bias (i.e., trials with adequate allocation  
55  
56 concealment, blinded outcome assessment, and intention to treat analysis) were compared to trials at high  
57  
58  
59  
60

1  
2  
3  
4 risk of bias. The effect of age was assessed by comparing trials including older participants (mean age >59  
5  
6 years) to trials including younger participants (mean age <60 years). The effect of type of exercise was  
7  
8 assessed by comparing trials using group exercises compared to trials using individual exercise. The effect  
9  
10 of duration of intervention was assessed by comparing trials with short duration of intervention to trials  
11  
12 with long duration of intervention splitting by the median time of duration. The effect of type of control  
13  
14 group was assessed by comparing trials using attention control to trials with waitlist controls and  
15  
16 comparing trials with exercise as add-on to medication to trials not using any medication. In addition, a  
17  
18 within-study comparison of low-dose exercise versus high-dose exercise in trials using different exercise  
19  
20 intensities was performed. The effect of co-morbid somatic disease was assessed by comparing the effect  
21  
22 estimates from trials including participants with depression compared to trials including participants with  
23  
24 depression in addition to a somatic disease. Publication bias was assessed by visual inspection of a funnel  
25  
26 plot and by Egger's test and if publication bias plausible Duval's and Tweedie's trim and fill procedure was  
27  
28 conducted.<sup>31</sup>  
29  
30  
31  
32  
33  
34  
35

36 We assessed and graded the evidence according to the grading of recommendations assessment,  
37  
38 development, and evaluation (GRADE) for high risk of bias, imprecision, indirectness, heterogeneity, and  
39  
40 publication bias.<sup>32</sup> Based on this assessment, the intervention was graded accordingly: 'high quality'- we are  
41  
42 very confident that the true effect lies close to that of the estimate of the effect; 'moderate quality'- we are  
43  
44 moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the  
45  
46 effect, but there is a possibility that it is substantially different; 'low quality'- our confidence in the effect  
47  
48 estimate is limited: the true effect may be substantially different from the estimate of the effect; 'very low  
49  
50 quality'- we have very little confidence in the effect estimate: the true effect is likely to be substantially  
51  
52 different from the estimate of the effect.<sup>33</sup>  
53  
54  
55  
56  
57  
58  
59  
60



### Deviations from our protocol

Post-hoc we included trials using the Chinese Classification of Mental Disorders (CCMD) as well as a few trials including participants classified as having 'minor depression'. The CCMD system closely adhere to the ICD and DSM systems and have been found highly compatible in field studies, so these studies were included.<sup>34</sup> A few trials included some participants classified as having 'minor depression' according to the trials chosen diagnostic system (e.g. DSM), and it is questionable if these participants have major depression. We therefore decided to include these trials but also to conduct a sub-group analysis exclusively including participants with major depression. To further explore heterogeneity, we post-hoc included sub-group analysis comparing intervention effects in inpatients and outpatients as well as an analysis according to trial size. Trials were divided into small or large trials using the median of total n included in the efficacy analysis. The effect of exercise capacity was post-hoc assessed by comparing trials with a high increase in maximal oxygen uptake (VO<sub>2</sub>max) with studies with lower increase in maximal oxygen uptake. Assessment of exercise capacity was based on the increase of VO<sub>2</sub>max in the intervention groups and trials were stratified to either high or low increase in exercise capacity by median. We did not conduct Trial Sequential Analysis based on a relative risk reduction of 30% of lack of remission as this was an implausible effect.

### Participant involvement

Depressed participants were not involved in this study.

### Results

#### Bibliographical search and trial characteristics

1  
2  
3  
4 The main bibliographical search was conducted the 26<sup>th</sup> of August, 2015 and the final updates were  
5  
6 conducted on the 20<sup>th</sup> of June, 2017. As illustrated in Figure S1, we identified 45 publications reporting the  
7  
8 effect of exercise on depressive symptoms in 35 randomised clinical trials.<sup>21;22;35-78</sup> Seven-teen trials were  
9  
10 conducted in Europe,<sup>21;22;40;49;52;53;55;61;65-68;74;75;77;79;80</sup> eight in the U.S.A.,<sup>38;39;43;45;60;64;76;81</sup>, six in Asia,<sup>47;69-73</sup> two  
11  
12 in Australia,<sup>54;58</sup> and two in South-America.<sup>56;63</sup> A total of 2,630 participants were randomised and 2,498  
13  
14 were included in the efficacy analysis of benefit. 10 trials included inpatients<sup>47;49;56;67;69-73;79</sup> and five trials  
15  
16 included participants with a mean age above 60 years.<sup>52;54;58;60;61</sup> No trials exclusively included participants  
17  
18 with comorbid somatic disease. Four trials reported the continuous outcome as mean change from baseline  
19  
20 in each group with a corresponding SD,<sup>39;53;65;68</sup> and one trial presented data as mean difference between  
21  
22 groups post-intervention.<sup>40</sup> The remaining trials reported post-scores in each group with corresponding SD.  
23  
24 Please see Table 1 for trial characteristics.  
25  
26  
27  
28  
29  
30

### 31 *Bias risk assessment*

32  
33  
34 Sequence generation was adequate in 15/35(43%), allocation concealment was adequate in 13/35 (37%)  
35  
36 trials, blinding of participants and trial personnel was adequate in 0/35 (0%), blinded outcome assessment  
37  
38 was performed in 16/35 (46%), low risk of bias in the 'incomplete outcome data' domain was found in  
39  
40 12/35 (34%) trials, selective outcome reporting domain was adequate in 31/35 (89%), for profit bias  
41  
42 domain was adequate in 19/35 (54%) and 25/35 (71%) were free of other bias. Accordingly, all trials were  
43  
44 at high risk of bias. Given the nature of the intervention, no trial had blinded participants or trial personnel,  
45  
46 however, two trials had low risk of bias in all other bias domains.<sup>22;54</sup> Five trials (16%) were sponsored by for  
47  
48 profit organisations: three trials were supported by pharmaceutical companies,<sup>53;79;82</sup> one trial by a  
49  
50 company producing fitness machines,<sup>45</sup> and one trial by an insurance company.<sup>21</sup> According to our a priori  
51  
52 defined criteria, 4/35 (11%) trials potentially had less risk of bias than the other trials at high risk of  
53  
54 bias.<sup>21;22;54;56</sup> Please see Table 2 for details on assessment of risks of bias.  
55  
56  
57  
58  
59  
60

## Primary outcomes

### The effect of exercise on depression severity

All included trials provided a continuous outcome on depression severity for the assessment of the exercise intervention encompassing 2,498/2,630 randomised participants (95%). The effect of intervention versus control was a standardised mean difference (SMD) of -0.66 (95% CI -0.86 to -0.46;  $P < 0.001$ ) (Figure 1.). This corresponds to an effect on the HAM-D<sub>17</sub> scale of -4.1 (95% CI -5.3 to -2.9) points.

### Missing data

Missing outcome analysis for depression as a continuous outcome did not markedly change the effect estimates. The least favourable outcome for the exercise intervention was the worse/best outcome analysis using +2 SD resulting in an effect estimate of -0.57 SMD (95% CI -0.78 to -0.36;  $P < 0.001$ ) (Table S1).

### Heterogeneity and subgroup analysis

The  $I^2$  was 81% suggesting substantial heterogeneity. Subgroup analysis revealed that the effect estimates for trials potentially having less risk of bias was -0.11 SMD (95% CI -0.41 to 0.18;  $P = 0.45$ ;  $I^2 = 62\%$ ) compared to that of the trials at high risk of bias -0.75 SMD (-0.98 to -0.52;  $p < 0.001$ ;  $I^2 = 81\%$ ) (test of subgroup difference,  $P < 0.001$ ). In addition, trials including 50 participants or less had a pooled estimate of -1.11 (-1.52 to -0.72;  $p < 0.001$ ;  $I^2 = 78\%$ ) compared to that of larger trials of -0.37 (-0.57 to -0.18;  $p < 0.001$ ;  $I^2 = 75\%$ ) (test of sub-group difference,  $P = 0.001$ ). Trials of short duration of intervention (less than 10 weeks) had a SMD of -0.92 (-1.09 to -0.74;  $p < 0.001$ ;  $I^2 = 14\%$ ) compared to trials with longer duration of intervention, -0.49 (-0.75 to -0.23;  $p < 0.001$ ;  $I^2 = 83\%$ ) (test of sub-group difference,  $P = 0.007$ ). Effect

1  
2  
3  
4 estimates from trials including participants with minor depression compared to trials exclusively including  
5  
6 participants with major depression did not differ (test of sub-group difference,  $P = 0.53$ ).  
7  
8

9  
10  
11 Four trials allocated 206 participants to different exercise intensities/doses.<sup>45;58;73;83</sup> Comparing the post-  
12  
13 intervention depression scores for participants allocated to either high intensity/high dose versus low  
14  
15 intensity/low dose exercise showed a difference of -0.40 SMD (95% CI -0.67 to -0.12;  $P=0.005$ ;  $I^2 = 0\%$ ) in  
16  
17 favour of high intensity/high dose exercise. As shown in Table 3, no other trial characteristic significantly  
18  
19 explained any of the observed heterogeneity. Please see Table S2 for trial characteristics used to explore  
20  
21 heterogeneity.  
22  
23  
24  
25  
26  
27  
28

### 29 *Trial Sequential Analysis and diversity adjusted required information size*

30  
31  
32 The diversity adjusted required information size for HAM-D<sub>17</sub> as a continuous outcome was calculated  
33  
34 based on our anticipated intervention effect of a minimal relevant difference of 3.0 HDRS points, a standard  
35  
36 deviation of 6.78 points, a risk of type I error of 0.05, a power of 90% and the observed diversity of 92% to  
37  
38 2610 participants. Only 14 trials reported results from HAM-D<sub>17</sub><sup>21;22;38;39;43;44;52;53;55;56;58;68;70;83</sup> with an accrued  
39  
40 1124 participants. As shown in Figure S2, the cumulative Z-curve just crossed the trial sequential  
41  
42 monitoring boundary for benefit. With the aforementioned settings, the pooled estimate is therefore less  
43  
44 likely to be a random finding due to lack of power or multiple testing if bias could be ignored. Post-hoc we  
45  
46 calculated the adjusted required information size for HAM-D<sub>17</sub> including all trials as shown in Figure S3. As  
47  
48 with the original analysis the Z-curve crossed the trial sequential monitoring boundary for benefit  
49  
50 supporting that the pooled estimate is less likely to represent a Type 1 error if bias could be ignored.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

### Bayes factor

Fourteen trials reported effect estimates using the HAM-D<sub>17</sub>.<sup>21;22;38;39;43;45;52;53;55;63;68;70;83;84</sup> Based on these trials, Bayes factor was calculated ( $\delta = -3.37$ ;  $SE_{\delta} = 0.96$ ;  $\mu_a = -3.0$ ) and was found to be 0.002, which is below the Bayes factor threshold for significance of 0.1, supporting the intervention effect if bias could be ignored.

### Publication bias

Inspection of the funnel-plot (not shown) suggested that small trials with small or no effect of exercise were missing (Figure S4). Egger's test supported the suspicion of publication bias,  $P < 0.00001$ . Using the Duval and Tweedie's trim and fill procedure, the estimate was reduced into -0.27 SMD (95% CI -0.50 to -0.05). This corresponds to an effect on the HAM-D<sub>17</sub> scale of -1.7 (95% CI -3.1 to -0.31) points.

### The effect of exercise on depression – lack of remission

Nineteen trials, randomising 1825 participants and including 1639 participants (90%) in final analysis reported remission as an outcome.<sup>21;22;38-40;43;45;47;49;53;54;56;60;61;65;68-70;72</sup> Remission post-intervention was defined in various ways: a post-intervention score on the HAM-D<sub>17</sub> less than 8 points,<sup>44;53;56;69;70</sup> not fulfilling the DSM criteria for depression *and* a HAM-D<sub>17</sub> less than 8 points,<sup>21;22;39</sup> not fulfilling the DSM criteria for depression,<sup>38;54;60</sup> a BDI score less than 9 points,<sup>43</sup> a BDI score less than 10 points,<sup>40</sup> a HAM-D<sub>17</sub> score less than 10 points,<sup>83</sup> a MADRS score less than 10 points,<sup>47</sup> a MADRS score less than 10 points *and* a 50% reduction in symptom score,<sup>65</sup> a 75% reduction in HAM-D<sub>24</sub>,<sup>72</sup> a HAM-D<sub>17</sub> score less than 11.28 points *and* a reduction in HAM-D<sub>17</sub> scores  $> 7.74$  points,<sup>68</sup> and one study used MADRS not specifying the cut-off for remission.<sup>49</sup> The RR for lack of remission was 0.78 (95% CI 0.68 to 0.90;  $P=0.0008$ ) in favour of the

1  
2  
3  
4 intervention using a random-effects analysis. The  $I^2$  was 69% suggesting substantial heterogeneity. The  
5  
6 forest plot for the intervention effect on lack of remission is illustrated in Figure S5.  
7  
8  
9

#### 10 11 *Missing data*

12  
13  
14 The scenario in least favour of the intervention was the 'poor' outcome analysis having an effect estimate  
15  
16 of RR 0.88 (95% CI 0.83 to 0.94)  $P=0.0002$ ;  $I^2 = 69\%$ . As shown in Table S1, the remaining scenarios did not  
17  
18 substantially differ from the main analysis.  
19

#### 20 21 22 *Heterogeneity and subgroup analysis*

23  
24  
25  $I^2$  was 69% for the outcome lack of remission suggesting substantial heterogeneity. For this outcome, only  
26  
27 two trials<sup>22;84</sup> were considered as trials potentially having less risk of bias than the other trials at high risk of  
28  
29 bias. The RR of these two trials was 0.95 (95% CI 0.74 to 1.23;  $P=0.78$ ) compared to 0.77 (96% CI 0.64 to  
30  
31 0.92;  $P=0.003$ ) for trials at high risk of bias, test of subgroup difference,  $P=0.19$ ). Trials including 52  
32  
33 participants or less in their final analysis had a RR 0.62 (95% CI 0.50 to 0.76;  $P<0.001$ ;  $I^2 = 45\%$ ) compared to  
34  
35 0.95 (95% CI 0.80 to 1.12;  $P=0.52$ ;  $I^2 = 68\%$ ) for larger trials (test of sub-group difference,  $P=0.002$ ). Also,  
36  
37 trials with a duration of less than 10 weeks had a RR of 0.63 (95% CI 0.51 to 0.77;  $P<0.001$ ;  $I^2 = 40\%$ )  
38  
39 compared to 0.93 (95% CI 0.78 to 1.10;  $P=0.39$ ;  $I^2 = 69\%$ ) for trials of a longer duration (test of sub-group  
40  
41 difference,  $P=0.004$ ). As shown in Table S3, no other trial characteristic significantly explained any of the  
42  
43 observed heterogeneity. Please see Table S2 for trial characteristics used to explore heterogeneity.  
44  
45  
46  
47  
48  
49

#### 50 51 *Trial Sequential Analysis and diversity adjusted required information size*

52  
53 The diversity adjusted required information size for lack of remission was calculated based on our observed  
54  
55 diversity of 74%, a proportion in the control group with lack of remission of 66%, an anticipated  
56  
57  
58  
59  
60

1  
2  
3  
4 intervention effect of 15% relative risk reduction, a risk of type I error of 0.05% and a power of 90%. As  
5  
6 shown in Figure S6, the cumulative Z curve just crossed the trial sequential monitoring boundary for  
7  
8 benefit. With the aforementioned settings, the pooled estimate is therefore less likely to be a random  
9  
10 finding due to lack of power or multiple testing if bias could be ignored.  
11  
12  
13

### 14 15 16 17 18 *Bayes factor*

19  
20 Bayes factor was calculate based on the observed relative risk of remission, the associated standard error,  
21  
22 and an anticipated intervention effect of relative increase in number of participants with remission by 15%  
23  
24 ( $\delta = -0.248$ ;  $SE_{\delta} = 0.08$ ;  $\mu_{\delta} = -0.163$ ). Bayes factor was 0.02, which is below the Bayes factor threshold for  
25  
26 significance of 0.1.  
27  
28  
29  
30  
31  
32

### 33 *Publication bias*

34  
35 Inspection of the funnel-plot (not shown) suggested that small trials with small or no effect of exercise  
36  
37 were missing. Egger's test supported the suspicion of publication bias,  $P=0.002$ . Imputing theoretically  
38  
39 missing studies by the Duval and Tweedie's trim and fill procedure, reduced the estimate of intervention  
40  
41 effect into a relative risk reduction of 0.93 (95% CI 0.79 to 1.11).  
42  
43  
44  
45  
46  
47

### 48 **The effect of exercise on serious adverse events**

49  
50 Serious adverse events (i.e., death or suicide attempts) were reported in only three trials.<sup>21,22,58</sup> In these  
51  
52 trials, one suicide attempt<sup>22</sup> and one death by suicide<sup>21</sup> were recorded in the intervention groups. The RR  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 for death or suicide in the two trials was 2.21 (95% CI 0.24 to 20.21; P=0.48; I<sup>2</sup> = 0%) as illustrated in Figure  
5  
6 S7.  
7  
8

#### 9 10 11 *Missing data*

12  
13  
14 Missing outcome analysis for 'serious adverse events' varied according to missing data scenario: poor  
15  
16 outcome analysis relative risk, 0.92 (95% CI 0.37 to 2.30; P=0.86; I<sup>2</sup> = 60.0%), good outcome analysis, 2.19  
17  
18 (95% CI 0.23 to 20.76; P=0.50; I<sup>2</sup> = 0.0%), best/worst outcome analysis – 0.08 (95% CI 0.02 to 0.34; P=0.001;  
19  
20 I<sup>2</sup> = 5.4%), worst/best outcome analysis 19.17 (95% CI 2.64 to 139.2; P=0.004; I<sup>2</sup> = 0.0%).  
21  
22  
23  
24  
25  
26

#### 27 *Trial Sequential Analysis and Bayes analysis*

28  
29 We decided not to conduct Trial Sequential Analysis or Bayes analysis due to too sparse data.  
30  
31  
32  
33  
34

#### 35 *Publication bias*

36  
37  
38 Only 3/35 trials reported on this outcome and no formal assessment for publication bias was made.  
39  
40 However, the lack of reporting in the vast majority of trials suggest risk publication bias.  
41  
42  
43  
44

#### 45 **Secondary outcomes**

##### 46 47 *The effect of exercise on quality of life*

48  
49 Nine trials randomising 827 participants reported on quality of life,<sup>21;22;38;40;56;60;71;76;85</sup> observing that  
50  
51 participants allocated to exercise did not have significantly better quality of life (SMD 0.40; 95% CI -0.03 to  
52  
53 0.83; P=0.07). The I<sup>2</sup> was 88% showing substantial heterogeneity (Figure S8).  
54  
55  
56  
57  
58  
59  
60



### *Non-serious adverse events*

Non-serious adverse events were reported in only ten trials.<sup>21;22;39;56;58;60;65;67;68;75</sup> Five trials reported on musculoskeletal adverse events without conducting formal tests<sup>58;60;65;67;68</sup> and four trials reported on number of participants with high depression scores post-intervention compared to baseline assessment.<sup>21;22;65;68</sup> The RR for increased severity of depression in patients allocated to exercise post-intervention was 0.83 (95% CI 0.40 to 1.70; P=0.60; I<sup>2</sup> = 0.0%).

### *The effect of exercise on depression beyond the duration of the intervention*

Assessment of depression beyond the intervention was conducted in seven trials,<sup>21;38;40;52;60;63;86</sup> with a median duration between end of intervention and assessment of depression of 6 months (range 5 to 23.5 months). The SMD between the intervention group and the control group using a random effects analysis was -0.10 (95% CI -0.28 to 0.09; P=0.31; I<sup>2</sup> = 19.5%). The I<sup>2</sup> for this estimate was 19.5% suggesting low heterogeneity (See Figure S9).

Remission beyond the intervention was assessed in five trials,<sup>21;38-40;54</sup> and the relative risk of lack of remission was 0.95 (95% CI 0.82 to 1.11; P=0.53) with an I<sup>2</sup> of 0.0% (See Figure S10).

### *GRADE assessments*

The GRADE assessments are presented in Table 4, and quality of evidence for both primary and secondary outcomes was very low or low.

### Additional analysis

Four studies reported change in scores from baseline with corresponding SD's, and one study reported mean difference between groups post-intervention. Comparing the effect size of these five studies with the remaining did not seem to explain part of the heterogeneity ( $p = 0.23$ ).

### Discussion

Thirty-five clinical trials allocating more than 2498 participants diagnosed with depression according to validated diagnostic instruments were included in the present systematic review. Pooled estimates suggested moderate antidepressant effect assessed both as a continuous outcome and as lack of remission. Due to risk of bias, inconsistency of effect estimates, and publication bias we have, however, very little confidence in these effect estimates. Subgroup analyses exploring reasons for the heterogeneity found that trials potentially having less risk of bias than other trials at high risk of bias had no effect of exercise on depression. Furthermore, duration of intervention and trial size were inversely associated with effect estimates. Exercise did not improve quality of life or depression or remission after the intervention. Serious adverse event or adverse events were reported inconsistently and only by a few trials not permitting firm conclusions regarding these outcomes.

### *Strengths and limitations*

The strengths of this systematic review are that it is based on the published protocol, a comprehensive search strategy, and the inclusion of patient centered outcomes such as quality of life as well as adverse events. Also, to avoid spurious finding from repeated testing, Trial Sequential Analysis and Bayes analysis were undertaken and these analyses did not suggest that the pooled estimates could be reduced to random errors for effect on depression severity or no remission. Neither Trial Sequential Analysis nor Bayes

1  
2  
3  
4 factor analysis are, however, able to wash of spurious effects induced by bias, fraud or other reasons.<sup>26;29;87-</sup>

5  
6<sup>89</sup> Had we restricted the Trial Sequential Analysis to trials of potentially lower risk of bias, the number of  
7  
8 trials and participants would be limited and we had seen evidence far from crossing any boundaries for  
9  
10 benefit, harms, or futility. The conclusions for serious adverse events and adverse events were associated  
11  
12 with wide confidence intervals due to lack of data and firm conclusions for these outcomes are presently  
13  
14 not available.

15  
16  
17  
18  
19  
20 The number of trials with adequate allocation concealment was 37% in the current systematic review  
21  
22 compared to only 15.1% in trials assessing non-drug interventions for depression.<sup>90</sup> Blinded outcome  
23  
24 assessment was performed in 46% of the included trials compared to 44% in non-drug antidepressant trials  
25  
26 in general.<sup>90</sup> The incomplete outcome bias domain was adequate in 34% of our included trials compared to  
27  
28 32.9% of antidepressant non-drug trials in general.<sup>90</sup> Compared to non-drug trials assessing interventions  
29  
30 for participants with depression, the included exercise trials have more bias domains with low risk of bias.  
31  
32 However, all our included trials were at high risk of bias. Two trials had low risk of bias for all bias domains  
33  
34 except for blinding of participants and trial personnel, and four trials fulfilled our criteria for trials at  
35  
36 potentially less risk of bias than the rest of the trials with at risk of bias. Despite a search strategy including  
37  
38 bibliographical databases and trials from China and South-America, the vast majority of included trials were  
39  
40 conducted in north America and western Europe, which is comparable to the geographical distribution of  
41  
42 non-drug trials in general<sup>90</sup> limiting the applicability to other geographic regions.

43  
44  
45  
46  
47  
48  
49 All outcomes for the primary analysis reflect depression severity, however, the different psychometrics may  
50  
51 represent different aspects of depression not reflected in the pooled estimate. An in-depth discussion of  
52  
53 the included assessment scales is beyond the scope of this review, but in the current systematic review we  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 found no significant differences of effect estimates from trials using HAM-D<sub>17</sub> compared to trials using  
5  
6 other assessment scales (data not shown).  
7  
8  
9

### 10 11 *The effect of exercise on depression*

12  
13  
14 Our present results are similar to the latest Cochrane review by Cooney et al. (2013)<sup>24</sup> who found a  
15 moderate effect of exercise on depressive symptoms (-0.62 SMD) when including all trials and no effect  
16 when restricting the analysis to trials with less risk of bias (-0.18 SMD). The Cochrane review did find  
17 evidence of a small antidepressant effect beyond the intervention, which we could not confirm in our  
18 present systematic review. Bridle et al. (2012)<sup>13</sup> included 9 trials allocating old (> 60 years) participants with  
19 depression to exercise interventions versus control interventions. Restricting the analysis to four trials at  
20 lower risk of bias they found small to moderate effect estimates (SMD -0.34) in favour of exercise. The  
21 studies by Cooney et al.<sup>24</sup> and Bridle et al.<sup>13</sup> both included trials allocating participants with depressive  
22 symptoms and not necessarily diagnosed using a validated diagnostic system, potentially explaining the  
23 differences in the effect sizes. However, in our present systematic review the estimate for four trials at  
24 potential less risk of bias than the remaining trials was -0.11 SMD and in the Cooney study the effect  
25 estimate for eight trials with lower risk of bias was -0.18 SMD<sup>24</sup> compared to -0.34 in the study by Bridle et  
26 al.<sup>13</sup> Meta-analysis of randomised clinical trials assessing the effects of exercise for depression consistently  
27 finds positive effects, however, when restricting the analysis to trials with less risk of bias the pooled effect  
28 sizes becomes very small or negligible. Meta-analysis examining the effect of exercise beyond the  
29 intervention also finds no or small effects of exercise. In the process of interpretation of effect estimates in  
30 the current research field, it is important to recognise that effect estimates from trials with non-blinded  
31 outcome assessment are at high risk of bias as reported by Savovic et al.<sup>91</sup> Sixteen of 35 trials in the current  
32 systematic review did not use blinded outcome assessment. In contradiction to the current systematic  
33 review, a recent meta-analysis by Schuch et al.<sup>12</sup> concluded that “exercise has a large and significant  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 antidepressant effect in people with depression.....Our data strongly support the claim that exercise is an  
5  
6 evidence-based treatment for depression". This statement was based on a meta-analysis of 25 randomised  
7  
8 clinical trials including participants with depression or depressive symptoms to exercise or control  
9  
10 conditions and excluding trials using any form of active control group. Surprisingly, the authors found that  
11  
12 adjusting for publication bias using the Trim and Fill procedure<sup>31</sup> the estimate *increased* from a SMD of  
13  
14 0.98 to 1.11. The effect in SMD in included studies ranged from -0.23 to 4.56 representing considerable  
15  
16 heterogeneity.<sup>12</sup> The authors classified four trials as having lower risk of bias using the same criteria as in  
17  
18 our systematic review and 21 trials as having high risk of bias. This illustrates some of the challenges in  
19  
20 meta-analysis of exercise and depression: the large heterogeneity driven by small studies inflating the  
21  
22 effects of random-effects analysis,<sup>92</sup> the misconception that we can restrict our analysis to statistics and not  
23  
24 consider the evident effect of bias.<sup>23;91</sup> Compared to our previous review,<sup>10</sup> we now included 35 trials  
25  
26 including 2498 participants versus previously 13 trials and 687 participants. It may seem as a paradox that  
27  
28 this large increase in data has not provided us with a similar increase in certainty of conclusions reflected  
29  
30 by heterogeneity of trial results as well as our conclusions from the systematic reviews. The increase in  
31  
32 available data is, however, primarily provided by small trials at high risk of bias introducing exaggerated  
33  
34 effect estimates. In the current systematic review, we included four trials with 530 participants at lower risk  
35  
36 of bias compared to three trials with 239 participants in our previous review, reflecting that only a small  
37  
38 part of the additional data comes from trials at lower risk of bias. The continuous increase in data  
39  
40 associated with high risk of bias will not provide patients, clinicians or policymakers with adequate  
41  
42 information and represents an unethical enrollment of trial participants and waste of resources.<sup>93-99</sup> We  
43  
44 therefore recommend that future systematic reviews and meta-analysis a priori should have a primary  
45  
46 outcome restricting effect analysis to larger trials with lower risk of bias and that any recommendations  
47  
48 regarding exercise interventions for participants with depression should be assessed with the GRADE  
49  
50 framework.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 The  $I^2$  of 81% and 69% for the primary outcomes indicate substantial evidence of heterogeneity of  
5  
6 intervention effects that is variation in effect estimates beyond chance. Part of this heterogeneity was  
7  
8 explained by bias and by trial size: trials at high risk of bias or small trials have very large effect estimates  
9  
10 compared to trials potentially at less risk of bias or larger trials. The funnel plots and Egger's test indicates  
11  
12 publication bias, however, the association between trial size and effect estimates could suggest that the  
13  
14 asymmetry in the funnel plots are due to small study bias rather than publication bias.<sup>100</sup> It could be argued  
15  
16 that both the delivery of exercise as well as the actual increase in fitness are fundamental to the  
17  
18 assessment of the antidepressant effects of exercise, and in line with our previous review we found  
19  
20 duration of intervention inversely associated with effect size.<sup>11</sup> Comparing different exercise intensities, we  
21  
22 did find a small effect of high intensity exercise compared to lower intensity exercise. However, assessing  
23  
24 delivered exercise expressed as increase in maximal oxygen uptake we could not reproduce this finding.  
25  
26 Future trials need to pay more attention to the dose of the intervention as well as compliance with  
27  
28 intervention.<sup>101</sup> We suggest using maximal oxygen uptake or 1 repetition maximum as the gold standards to  
29  
30 assess the received exercise. Several studies compare exercise to control interventions rather than wait-list  
31  
32 control to reduce the effect of non-specific effects, e.g. the DEMO trials and Mather et al.<sup>21;22;52</sup> Also, it  
33  
34 could be speculated that the effect of exercise would be harder to detect if participants also received  
35  
36 medical treatment in addition. The current systematic review could not confirm that the type of control  
37  
38 condition explained heterogeneity. The discussion of control group is important in non-drug trials: choosing  
39  
40 a waitlist control group the results potentially reflects non-specific effects, choosing an active control group  
41  
42 (e.g., relaxation exercise) the trial is potentially a comparison between two active treatments. However, in  
43  
44 the current systematic review we found no evidence that trials using an attention control group or exercise  
45  
46 as add-on to pharmacotherapy had significantly different effect estimates compared to other trials.  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 Our systematic review did not find indications of a positive effect on quality of life in participants with  
5 depression allocated to exercise interventions, which is in concordance with the review by Cooney et al.<sup>24</sup>  
6  
7  
8 Only 3/35 trials reported on serious adverse events, and we found no significant effects of exercise on risk  
9 of death or suicide attempt. No indication of increased severity of depression or other adverse events in  
10 participants allocated to exercise could be detected. However, data on adverse events was reported  
11 sporadically in a minority of trials and currently it is not possible to conclude on the risk of serious adverse  
12 events or adverse event for exercise interventions in participants with depression.  
13  
14  
15  
16  
17  
18  
19  
20  
21

### 22 *Conclusions*

23  
24 We have little confidence in the pooled effect estimates, especially because trials with less than high risk of  
25 bias produced significantly lower effect estimates, suggesting that exercise interventions only produce  
26 small or negligible antidepressant effects, depending on how much of the effect is caused by bias and how  
27 much is caused by the intervention. There was no effect of exercise on depression beyond the intervention  
28 itself. We found no effect on quality of life. There is currently no evidence in favour of exercise for patients  
29 with depression with a view to ameliorate depressive symptoms. Our systematic review did not evaluate  
30 possible beneficial effects of exercise on, e.g., metabolism or cardiovascular fitness,<sup>22,102</sup> and it is possible  
31 that exercise may have beneficial effects on these factors in patients diagnosed with depression.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

### 46 *Future perspectives*

47  
48 Despite the large number of published trials, further trials with more robust methodology seem still  
49 required to establish progress in this field. Also, additional trials from outside North-America and Europe  
50 may be required for results to be valid for patients in Asia, Africa and South-America. To further elaborate  
51 on the current findings, we recommend that future trials must include blinded outcome assessors and  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 outcomes assessing quality of life, metabolic effects, and long-term effects beyond the intervention. It is  
5  
6 also important that future trials systematically collect and report data on death, suicide events,  
7  
8 musculoskeletal injuries and other potential adverse effects in both the intervention group as well as in the  
9  
10 control group. Moreover, future trials ought to be designed according to the SPIRIT guidelines and reported  
11  
12 according to the CONSORT guidelines<sup>103;104</sup> and transparently report deidentified individual participant data  
13  
14 enabling individual participant data meta-analyses.<sup>105</sup>  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



## Funding

No funding was obtained for this systematic review.

## Competing interests

JK, CG, and MN have previously published two trials and a meta-analysis on this topic, which could introduce an academic bias in the current systematic review. We asked new authors (HS and CH) to be involved in the preparation of the protocol, trial selection and bias assessment. No support from any organisation was received for the submitted work; no financial relationship with any organisations that might have an interest in the submitted work in the previous three years; and apart from the above no other relationship or activities that could appear to have influenced the submitted work.

## Acknowledgement

We appreciate the help from Youling He for help with the Chinese Wanfang bibliographical database and translation of Chinese papers. We thank Janus C Jakobsen for assistance with the calculation of Bayes factor.

## Contributors

JK conceived the project, collected data, did the statistical analysis, analysed the data, drafted and revised the manuscript. He is guarantor. CH collected the data, analysed the data and revised the

1  
2  
3  
4 manuscript. HS conceived the project, collected data, analysed the data, and revised the  
5  
6 manuscript. CG conceived the project, analysed the data and revised the manuscript. MN  
7  
8 conceived the project, analysed the data, and revised the manuscript.  
9  
10

#### 11 12 13 14 15 **Data sharing statement**

16  
17  
18 All data used in this study are available in Figures and Tables. No other data were used.  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Reference List

- 1  
2  
3  
4  
5  
6  
7  
8 (1) Lepine JP, Gastpar M, Mendelwicz J, Tylee A. Depression in the community: The first pan-  
9 european study DEPRES (Depression Research in European Society). *Int Clin Psychopharmacol*  
10 1997; 12:19-29.
- 11  
12 (2) Pedersen CB, Mors O, Bertelsen A, *et al.* A comprehensive nationwide study of the incidence rate  
13 and lifetime risk for treated mental disorders. *JAMA Psychiatry* 2014; 71(5):573-581.
- 14  
15 (3) Ustun TB, Ayuso-Mateos JL, Chatterji S, *et al.* Global burden of depressive disorders in the year  
16 2000. *Br J Psychiatry* 2004; 184:386-392.
- 17  
18 (4) Kirsch I, Deacon B, Huedo-Medina T, *et al.* Initial severity and antidepressant benefits: A meta-  
19 analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008; 5[2], e45.  
20 doi:10.1371/journal.pmed.0050045.
- 21  
22 (5) Turner EH, Matthews AM, Linardatos E, *et al.* Selective Publication of Antidepressant Trials and  
23 Its Influence on Apparent Efficacy. *N Engl J Med* 2008; 358(3):252-260.
- 24  
25 (6) Jakobsen JC, Katakam KK, Schou A, *et al.* Hellmuth SG, Stallknecht SE, Leth-Moller K *et al.*  
26 Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder.  
27 A systematic review with meta-analysis and Trial Sequential Analysis. *BMC Psychiatry* 2017;  
28 17(1):58.
- 29  
30 (7) Jakobsen JC, Hansen JL, Storebo OJ, *et al.* The effects of cognitive therapy versus 'no intervention'  
31 for major depressive disorder. *PLoS One* 2011; 6(12):e28299.
- 32  
33 (8) Jakobsen JC, Hansen JL, Simonsen E, Gluud C. The effect of interpersonal psychotherapy and  
34 other psychodynamic therapies versus 'treatment as usual' in patients with major depressive  
35 disorder. *PLoS One* 2011; 6(4):e19044.
- 36  
37 (9) Jakobsen JC, Hansen JL, Simonsen E, Gluud C. The effect of adding psychodynamic therapy to  
38 antidepressants in patients with major depressive disorder. A systematic review of randomized  
39 clinical trials with meta-analyses and trial sequential analyses. *J Affect Disord* 2012; 137(1-3):4-14.
- 40  
41 (10) Cipriani A, Furukawa TA, Salanti G, *et al.* Comparative efficacy and acceptability of 12 new-  
42 generation antidepressants: a multiple-treatments meta-analysis. *Lancet* 2009; 373: 746-758.
- 43  
44 (11) Krogh J, Nordentoft M, Sterne J, Lawlor D. The effect of exercise in clinically depressed adults:  
45 Systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry* 2011;  
46 72:529-538.
- 47  
48 (12) Schuch FB, Vancampfort D, Richards J, *et al.* Exercise as a treatment for depression: A meta-  
49 analysis adjusting for publication bias. *J Psychiatr Res* 2016; 77:42-51.
- 50  
51 (13) Bridle C, Spanjers K, Patel S, *et al.* Effect of exercise on depression severity in older people:  
52 systematic review and meta-analysis of randomised controlled trials. *Br J Psychiatry* 2012;  
53 201(3):180-185.
- 54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (14) Herring MP, Puetz TW, O'Connor PJ, Dishman RK. Effect of exercise training on depressive  
5 symptoms among patients with a chronic illness: a systematic review and meta-analysis of  
6 randomized controlled trials. *Arch Intern Med* 2012; 172(2):101-111.  
7  
8 (15) World Health Organisation. International Statistical Classification of Diseases, 10th Revision (ICD-  
9 10). *Geneva, Switzerland* 1992.  
11  
12 (16) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 4th ed.  
13 *Washington DC* 1994.  
14  
15 (17) Beck AT, Steer RA, Brown GK. BDI-II manual. 2 ed. New York: Psychological Corporation, 1996.  
16  
17  
18 (18) Krogh J, Speyer H, Gluud C, Nordentoft M. Exercise for patients with major depression: A  
19 systematic review with meta-analysis and trial sequential analysis. *Syst Rev* 2015; 4(1):40.  
20  
21 (19) Spitzer RL, Endicott J, Williams JB. Research diagnostic criteria. *Arch Gen Psychiatry* 1979;  
22 36(12):1381-1383.  
23  
24 (20) ICH-secretariat. International Conference on Harmonisation (ICH) Harmonised Tripartite  
25 Guideline (Step 4). Guideline for Good Clinical Practice. *Geneva* 1996.  
26  
27 (21) Krogh J, Saltin B, Gluud C, Nordentoft M. The DEMO trial: A randomized, parallel-group,  
28 observer-blinded clinical trial of strength versus aerobic versus relaxation training for patients  
29 with mild to moderate depression. *J Clin Psychiatry* 2009; 70[6]: 790-800.  
30  
31 (22) Krogh J, Videbech P, Thomsen C, *et al.* DEMO-II trial. Aerobic Exercise Versus Stretching Exercise in  
32 Patients With Major Depression - A Randomised Clinical Trial. *PLoS One* 2012; 7(10): e48316.  
33 doi:10.1371/journal.pone.0048316.  
34  
35 (23) Higgins JPT GSe. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The  
36 Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org) The Cochrane  
37 Collaboration, 2009.  
38  
39 (24) Cooney GM, Dwan K, Greig CA, *et al.* Exercise for depression. *Cochrane Database Syst Rev* 2013;  
40 9:CD004366.  
41  
42 (25) Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analysis. *BMJ*  
43 2003; 327:557-560.  
44  
45 (26) Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm  
46 evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008; 61(1):64-75.  
47  
48 (27) Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information  
49 size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008; 61(8):763-  
50 769.  
51  
52 (28) Thorlund K, Walter SD, Johnston BC, *et al.* Pooling health-related quality of life outcomes in meta-  
53 analysis-a tutorial and review of methods for enhancing interpretability. *Res Synth Methods* 2011;  
54 2(3):188-203.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (29) Jakobsen JC, Wetterslev J, Winkel P, *et al.* Thresholds for statistical and clinical significance in  
5 systematic reviews with meta-analytic methods. *BMC Med Res Methodol* 2014; 14:120.  
6  
7 (30) Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published  
8 randomised controlled trials. *BMJ* 1999; 319(7211):670-674.  
9  
10 (31) Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for  
11 publication bias in meta-analysis. *Biometrics* 2000; 56(2):455-463.  
12  
13 (32) Guyatt GH, Oxman AD, Vist GE, *et al.* GRADE: an emerging consensus on rating quality of  
14 evidence and strength of recommendations. *BMJ* 2008; 336(7650):924-926.  
15  
16 (33) Andrews J, Guyatt G, Oxman AD, *et al.* GRADE guidelines: 14. Going from evidence to  
17 recommendations: the significance and presentation of recommendations. *J Clin Epidemiol* 2013;  
18 66(7):719-725.  
19  
20 (34) Zheng YP, Lin KM, Zhao JP, *et al.* Comparative study of diagnostic systems: Chinese Classification  
21 of Mental Disorders-Second Edition versus DSM-III-R. *Compr Psychiatry* 1994; 35(6):441-449.  
22  
23 (35) Babyak MA, Blumenthal JA, Herman S, *et al.* Exercise treatment for major depression:  
24 Maintenance of therapeutic benefit at 10 months. *Psychosom Med* 2000; 62:633-638.  
25  
26 (36) Baxter H, Winder R, Chalder M, *et al.* Physical activity as a treatment for depression: the TREAD  
27 randomised trial protocol. *Trials* 2010; 11:105.  
28  
29 (37) Benson M, Blumenthal J, Babyak M, *et al.* Exercise fails to improve neurocognition in depressed  
30 middle-aged and older adults. *Med Sci Sports Exerc* 2008; 40[7]: 1344-1352.  
31  
32 (38) Blumenthal J, Babyak MA, Moore KA, *et al.* Effects of exercise training on older patients with  
33 major depression. *Arch Int Med* 1999; 159:2349-2356.  
34  
35 (39) Blumenthal J, Babyak M, Doraiswamy P, *et al.* Exercise and pharmacotherapy in the treatment of  
36 major depressive disorder. *Psychosom Med* 2007; 69: 587-596.  
37  
38 (40) Chalder M, Wiles NJ, Campbell J, *et al.* Facilitated physical activity as a treatment for depressed  
39 adults: randomised controlled trial. *BMJ* 2012; 344:e2758.  
40  
41 (41) Chalder M, Wiles NJ, Campbell J, *et al.* A pragmatic randomised controlled trial to evaluate the  
42 cost-effectiveness of a physical activity intervention as a treatment for depression: the treating  
43 depression with physical activity (TREAD) trial. *Health Technol Assess* 2012; 16(10):1-iv.  
44  
45 (42) Combs K, Smith PJ, Sherwood A, *et al.* Impact of sleep complaints and depression outcomes  
46 among participants in the standard medical intervention and long-term exercise study of exercise  
47 and pharmacotherapy for depression. *J Nerv Ment Dis* 2014;202(2):167-71  
48  
49 (43) Doyne EJ, Ossip-Klein DJ, Bowman ED, *et al.* Running versus weight lifting in the treatment of  
50 depression. *J Consult Clin Psychol* 1987; 55(5):748-754.  
51  
52 (44) Dunn AL, Trivedi MH, Kampert JB, *et al.* The DOSE study: a clinical trial to examine efficacy and  
53 dose response of exercise as treatment for depression. *Control Clin Trials* 2002; 23(5):584-603.  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (45) Dunn AL, Trivedi MH, Kaplan K, *et al.* Exercise treatment for depression. Efficacy and dose  
5 response. *Am J Prev Med* 2005; 28: 1-8.  
6  
7 (46) Herman S, Blumenthal JA, Babyak M, *et al.* Exercise therapy for depression in middle-aged and  
8 older adults: predictors of early dropout and treatment failure. *Health psychol* 2002; 21:553-563.  
9  
10 (47) Ho CWH, Chan SC, Wong JS, *et al.* Effect of aerobic exercise training on chinese population with  
11 mild to moderate depression in Hong Kong.  
12  
13 (48) Hoffman BM, Babyak MA, Craighead WE, Sherwood A, Doraiswamy PM, Coons MJ *et al.* Exercise  
14 and pharmacotherapy in patients with major depression: one-year follow-up of the SMILE study.  
15 *Psychosom Med* 2011; 73(2):127-133.  
16  
17 (49) Kerling A, Tegtbur U, Gutzlaff E, Kuck M, Borchert L, Ates Z *et al.* Effects of adjunctive exercise on  
18 physiological and psychological parameters in depression: A randomized pilot trial. *Rehabil Res*  
19 *Pract* 2014;2014:627376  
20  
21 (50) Khatri P, Blumenthal J, Babyak M, *et al.* Effects of exercise training on cognitive functioning  
22 among depressed older men and women. *J Aging Phys Act* 2001; 9: 43-57.  
23  
24 (51) Krogh J, Petersen L, Timmermann M, *et al.* Design paper: The DEMO trial: A randomized, parallel-  
25 group, observer-blinded clinical trial of aerobic versus non-aerobic versus relaxation training for  
26 patients with light to moderate depression. *Contemp Clin Trials* 2007; 28:79-89.  
27  
28 (52) Mather AS, Rodriguez C, Guthrie MF, *et al.* Effects of exercise on depressive symptoms in older  
29 adults with poorly responsive depressive disorder: randomised controlled trial. *Br J Psychiatry*  
30 2002; 180:411-415.  
31  
32 (53) Mota-Pereira J, Silverio J, Carvalho S, *et al.* Moderate exercise improves depression parameters in  
33 treatment-resistant patients with major depressive disorder. *J Psychiatr Res* 2011; 45(8):1005-  
34 1011.  
35  
36 (54) Pfaff JJ, Alfonso H, Newton RU, *et al.* ACTIVEDEP: a randomised, controlled trial of a home-based  
37 exercise intervention to alleviate depression in middle-aged and older adults. *Br J Sports Med*  
38 2014; 48(3):226-232.  
39  
40 (55) Pihu A, Sorba M, Hardoy M, *et al.* Efficacy of physical activity in the adjunctive treatment of major  
41 depressive disorders: preliminary results. *Clin Pract Epidemio Ment Health* 2007;3:8.  
42  
43 (56) Schuch FB, Vasconcelos-Moreno MP, Borowsky C, Fleck MP. Exercise and severe depression:  
44 preliminary results of an add-on study. *J Affect Disord* 2011; 133(3):615-618.  
45  
46 (57) Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of the effect of exercise on  
47 sleep. *Sleep* 1997; 20(2):95-101.  
48  
49 (58) Singh NA, Stavrinou TM, Scarbek Y, *et al.* A randomized controlled trial of high versus low  
50 intensity weight training versus general practitioner care for clinical depression in older adults. *J*  
51 *Gerontol A Biol Sci Med Sci* 2005; 60(6):768-776.  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (59) Singh NA, Clements KM, Singh MA. The efficacy of exercise as a long-term antidepressant in  
5 elderly subjects: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci* 2001; 56(8):M497-  
6 M504.  
7  
8 (60) Singh N, Clements K, Fiatarone M. A randomized controlled trial of progressive resistance training  
9 in depressed elders. *Journal of Gerontology* 1997; 52:M27-M35.  
10  
11 (61) Toni G, Belvederi MM, Zanetidou S, *et al.* Physical exercise for late life Major Depression: The  
12 SEEDS study. *Eur J Prev Cardiol* 2014; 21:S107.  
13  
14 (62) Veale D, Le FK, Pantelis C, *et al.* Aerobic exercise in the adjunctive treatment of depression: a  
15 randomized controlled trial. *J R Soc Med* 1992; 85(9):541-544.  
16  
17 (63) Vieira JLL, Porcu M, da Rocha PGM. The regular physical exercise's practice as complementary  
18 therapy to the treatment of depressed women. *J Bras Psiquiatr* 2007;56(1):23-8.  
19  
20 (64) Epstein D. Reno: University of Nevada: Aerobic activity versus group cognitive therapy: An  
21 evaluative study of contrasting interventions for the alleviation of clinical depression  
22 (*dissertation*), 1986.  
23  
24 (65) Danielsson L, Papoulias I, Petersson E, *et al.* Exercise or basic body awareness therapy as add-on  
25 treatment for major depression: A controlled study. *J Affect Disord* 2014; 168:98-106.  
26  
27 (66) Carneiro L, Fonseca A, Vieira-Coelho M, *et al.* Effects of structured exercise and pharmacotherapy  
28 vs pharmacotherapy for adults with depressive symptoms: A randomized clinical trial. *J Psychiatr*  
29 *Res* 2015;(71):48-55.  
30  
31 (67) Legrand FD, Neff EM. Efficacy of exercise as an adjunct treatment for clinically depressed  
32 inpatients during the initial stages of antidepressant pharmacotherapy: An open randomized  
33 controlled trial. *J Affect Disord* 2016; 191:139-144.  
34  
35 (68) Doose M, Ziegenbein M, Hoos O, *et al.* Self-selected intensity exercise in the treatment of major  
36 depression: A pragmatic RCT. *Int J Psychiatry Clin Pract* 2015; 19(4):266-275.  
37  
38 (69) Salehi I, Hosseini SM, Haghighi M, *et al.* Electroconvulsive therapy (ECT) and aerobic exercise  
39 training (AET) increased plasma BDNF and ameliorated depressive symptoms in patients suffering  
40 from major depressive disorder. *J Psychiatr Res* 2016; 76:1-8.  
41  
42 (70) Huipeng F, Xiaohui Z. Curative effect and observation of exercise on patients with depression.  
43 *Chin Med Pharm* 2013; 3:187-188.  
44  
45 (71) Jinchun W, Bo Y, Rong X, *et al.* Effect of exercise therapy in improving social function and quality  
46 of life in patients with depression. *Nurs J Chin PLA* 2015; 32(21):24.  
47  
48 (72) Guifeng S. The role of setting - up exercise on rehabilitation in patients with depressive episode.  
49 *Chin J Health Psychol* 2015; 23:184-186.  
50  
51 (73) Fang D, Min R, Ai-qin W. The effect of aerobic exercise with different intensity and frequency on  
52 depressive symptoms. *Chin J Behav Med & Brain Sci* 2013; 22:830-832.  
53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (74) Olson RL, Brush CJ, Ehmann PJ, *et al.* A randomized trial of aerobic exercise on cognitive control in  
5 major depression. *Clin Neurophysiol* 2017; 128(6):903-913.  
6  
7 (75) Euteneuer F, Dannehl K, Del RA, *et al.* Immunological effects of behavioral activation with  
8 exercise in major depression: an exploratory randomized controlled trial. *Transl Psychiatry* 2017;  
9 7(5):e1132.  
10  
11 (76) Patten C, Bronars C, Douglas K, *et al.* Supervised, Vigorous Intensity Exercise Intervention for  
12 Depressed Female Smokers: A Pilot Study. *Nicotine Tob Res* 2017; 19:77-86.  
13  
14 (77) Pentecost C, Farrand P, Greaves CJ, *et al.* Combining behavioural activation with physical activity  
15 promotion for adults with depression: findings of a parallel-group pilot randomised controlled  
16 trial (BACPac). *Trials* 2015; 16:367.  
17  
18 (78) Farrand P, Pentecost C, Greaves C, *et al.* A written self-help intervention for depressed adults  
19 comparing behavioural activation combined with physical activity promotion with a self-help  
20 intervention based upon behavioural activation alone: study protocol for a parallel group pilot  
21 randomised controlled trial (BACPac). *Trials* 2014; 15:196.  
22  
23 (79) Martinsen EW, Medhus A, Sandvik L. Effects of aerobic exercise on depression: A controlled  
24 study. *BMJ* 1985; 291:109.  
25  
26 (80) Veale D, Fevre K, Pantis C, *et al.* Aerobic exercise in the adjunctive treatment of depression: A  
27 randomised controlled trial. *J R Soc Med* 1992; 85:541-544.  
28  
29 (81) Ossip-Klein DJ, Doyne EJ, Bowman ED, *et al.* Effects of running or weight lifting on self-concept in  
30 clinically depressed women. *J Consult Clin Psychol* 1989; 57(1):158-161.  
31  
32 (82) Blumenthal JA, Babyak MA, Moore KA, *et al.* Effects of exercise training on older patients with  
33 major depression. *Arch Intern Med* 1999; 159(19):2349-2356.  
34  
35 (83) Belvederi MM, Amore M, Menchetti M, *et al.* Physical exercise for late-life major depression. *Br J*  
36 *Psychiatry* 2015; 207(3):235-242.  
37  
38 (84) Schuch FB, Vasconcelos-Moreno MP, Borowsky C, *et al.* Exercise and severe major depression:  
39 effect on symptom severity and quality of life at discharge in an inpatient cohort. *J Psychiatr Res*  
40 2015; 61:25-32.  
41  
42 (85) Carta MG, Hardoy MC, Pulu A, *et al.* Improving physical quality of life with group physical activity  
43 in the adjunctive treatment of major depressive disorder. *Clin Pract Epidemiol Ment Health* 2008;  
44 4:1.doi: 10.1186/1745-0179-4-1.  
45  
46 (86) Klein MH, Greist JH, Gurman AS, *et al.* A comparative outcome study of group psychotherapy vs.  
47 exercise treatments for depression. *Int J Ment Health* 1985; 13:148-177.  
48  
49 (87) Roberts I, Ker K, Edwards P, *et al.* The knowledge system underpinning healthcare is not fit for  
50 purpose and must change. *BMJ* 2015; 350:h2463.  
51  
52 (88) Gluud C, Jakobsen J, Imberger G, *et al.* Re: The knowledge system underpinning healthcare is not  
53 fit for purpose and must change - Responses to the opposing viewpoints of Roberts and  
54 colleagues and Tovey and colleagues. *BMJ* 2015; 350(h2463).  
55  
56  
57  
58  
59  
60



- 1  
2  
3  
4 (89) Thorlund K, Engstrøm J, Wetterslev J, *et al.* User manual for Trial Sequential Analysis (TSA).  
5 *Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen, Denmark P1-115,*  
6 *Available from <https://www.ctu.dk/tsa> 2011.*  
7  
8 (90) Krogh J, Hjorthoj CR, Jakobsen JC, *et al.* DEPEROR: Risks of systematic errors in drug and non-  
9 drug randomized clinical trials assessing intervention effects in patients with unipolar depression.  
10 *J Affect Disord* 2015; 179:121-127.  
11  
12 (91) Savovic J, Jones HE, Altman DG, *et al.* Influence of reported study design characteristics on  
13 intervention effect estimates from randomized, controlled trials. *Ann Intern Med* 2012;  
14 157(6):429-438.  
15  
16 (92) Poole C, Greenland S. Random-effects meta-analyses are not always conservative. *Am J Epidemiol*  
17 1999; 150(5):469-475.  
18  
19 (93) Ioannidis JP, Greenland S, Hlatky MA, *et al.* Increasing value and reducing waste in research  
20 design, conduct, and analysis. *Lancet* 2014; 383(9912):166-175.  
21  
22 (94) Chan AW, Song F, Vickers A, *et al.* Increasing value and reducing waste: addressing inaccessible  
23 research. *Lancet* 2014; 383(9913):257-266.  
24  
25 (95) Moher D, Glasziou P, Chalmers I, *et al.* Increasing value and reducing waste in biomedical  
26 research: who's listening? *Lancet* 2016; 387(10027):1573-1586.  
27  
28 (96) Glasziou P, Altman DG, Bossuyt P, *et al.* Reducing waste from incomplete or unusable reports of  
29 biomedical research. *Lancet* 2014; 383(9913):267-276.  
30  
31 (97) Chalmers I, Bracken MB, Djulbegovic B, *et al.* How to increase value and reduce waste when  
32 research priorities are set. *Lancet* 2014; 383(9912):156-165.  
33  
34 (98) Al-Shahi SR, Beller E, Kagan J, *et al.* Increasing value and reducing waste in biomedical research  
35 regulation and management. *Lancet* 2014; 383(9912):176-185.  
36  
37 (99) Garattini S, Jakobsen JC, Wetterslev J, *et al.* Evidence-based clinical practice: Overview of threats  
38 to the validity of evidence and how to minimise them. *Eur J Intern Med* 2016; 32:13-21.  
39  
40 (100) Sterne JA, Sutton AJ, Ioannidis JP, *et al.* Recommendations for examining and interpreting funnel  
41 plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011; 343:d4002.  
42  
43 (101) Slade SC, Dionne CE, Underwood M, Buchbinder R. Consensus on Exercise Reporting Template  
44 (CERT): Explanation and Elaboration Statement. *Br J Sports Med* 2016.  
45  
46 (102) Stubbs B, Rosenbaum S, Vancampfort D, *et al.* Exercise improves cardiorespiratory fitness in  
47 people with depression: A meta-analysis of randomized control trials. *J Affect Disord* 2016;  
48 190:249-253.  
49  
50 (103) Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting  
51 parallel group randomized trials. *Ann Intern Med* 2010; 152(11):726-732.  
52  
53 (104) Chan AW, Tetzlaff JM, Gotzsche PC, *et al.* SPIRIT 2013 explanation and elaboration: guidance for  
54 protocols of clinical trials. *BMJ* 2013; 346:e7586.  
55  
56  
57  
58  
59  
60

- 1  
2  
3  
4 (105) Skoog M, Saarimäki J, luud C, heinin M, rlandsson K, Aamdal S. Transparency and registration in  
5 clinical research in the Nordic countries. *Nordic Trial Alliance, NordForsk* 2015; 1-108  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Figure legends**

**Figure 1.** Effect of exercise on depression severity in patients diagnosed with depression

For peer review only

**Table 1.** Characteristics of trials assessing exercise for patients diagnosed with depression

Author, first Country of origin	Participants	Severity of depression at baseline	N at baseline (included in trial efficacy analysis)	Type of intervention	Frequency	Duration
Klein 1985 USA	Outpatients Mean age: 30 (SD 7) 72% female	SCL-D: 2.4 (SD 1)	50 (22)	<i>Aerobic exercise:</i> Supervised individual running. <i>Control group:</i> Supervised meditation in groups	2 sessions per week  Control group: 1 session per week	12 weeks
Martinsen 1985 Norway	Inpatients Mean age: 40 (range 17-60) Distribution of sex not reported	BDI: 28.0 (SD 9)	49(43)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Occupational therapy.	3 sessions per week  Control group: 3 sessions per week	9 weeks
Epstein 1986 USA	Outpatients Mean age: 39 (range 24 to 60) (NR) % female	BDI: 23.4 (SD 7)	21 (17)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Waitlist control.	3 sessions per week	8 weeks
Doyne 1987 USA	Outpatients Mean age: 29 (SD 4) 100 % female	HAM-D <sub>17</sub> : 13.0 (SD 7)	52 (25)	<i>Aerobic exercise OR weightlifting:</i> Supervised individual exercise. <i>Control group:</i> Waiting list.	4 sessions per week	8 weeks
Veale 1992 UK	Outpatients Mean age: 35 (range 19-58) 64% female	BDI: 24.5 (SD 6)	83 (65)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Standard treatment from psychiatric services.	3 sessions per week	12 weeks
Singh 1997 USA	Outpatients Recruited from a register of volunteers Mean age: 71 (SD 1)	BDI: 19.9 (SD 2.3)	32 (32)	<i>Progressive resistance training:</i> Supervised group exercise. <i>Control group:</i> Attended seminars on health.	3 sessions per week  Control group: 2 sessions per week	10 weeks
Blumenthal 1999 USA	Outpatients Mean age: 57 (SD 7) 71.8% female	HAM-D <sub>17</sub> : Not reported	103 (103)	<i>Aerobic exercise:</i> Supervised exercise plus antidepressant medication (sertraline). <i>Control group:</i> Antidepressant medication (sertraline).	3 sessions per week	16 weeks
Mather 2002 UK	Outpatients Treatment resistant Mean age: 65 (range 53-91) 69% female	HAM-D <sub>17</sub> : 17.1 (SD 6)	86 (85)	<i>Mixed aerobic and non-aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Attended health seminars.	2 sessions per week  Control group: 2 seminars per week	10 weeks
Dunn 2005 USA	Outpatients Mean age: 36 (SD 6)	HAM-D <sub>17</sub> : 19.4 (SD 2)	80 (80)	<i>Aerobic exercise:</i> Individually supervised	Group (1) and (2): 3 sessions	12 weeks

		75% female			exercise with (1) low energy expenditure (EE) OR (2) high EE OR (3) low EE OR (4) high EE. <i>Control group:</i> Flexibility exercise.	per week Group (3) and (4): 5 sessions per week Control group: 3 sessions per week	
10	Singh 2005 Australia	Outpatients Mean age: 69 (SD 6) 55% female	HAM-D <sub>17</sub> : 18.9 (SD 4.2)	60 (54)	<i>Progressive resistance training (PRT):</i> (1)Low intensity PRT OR (2) high intensity PRT. <i>Control group:</i> Standard GP care.	Group (1) and (2): 3 sessions per week	8 weeks
16	Pilu 2007 Italy	Outpatients Treatment resistant Age between 40 and 60 100% female	HAM-D <sub>17</sub> : 19.7 (SD 6)	30 (30)	<i>Resistance exercise:</i> Supervised group sessions. <i>Control group:</i> Standard treatment.	2 sessions per week	32 weeks
21	Viera 2007 Brazil	Outpatients Mean age 43.66 (SD NR) 100% female	HAM-D <sub>21</sub> : 31.9 (SD 3)	18 (18)	<i>Aerobic exercise:</i> Supervised water aerobics. <i>Control group:</i> Standard GP care.	2 sessions per week	12 weeks
26	Blumenthal 2007 USA	Outpatients Mean age: 52 (SD 8) 75.8% female	HAM-D <sub>17</sub> : 16.7 (SD 4)	153 (153)	<i>Aerobic exercise:</i> (1) Supervised group exercise OR (2) home-based exercise. <i>Control group:</i> Placebo medication.	(1) and (2): 3 sessions per week	16 weeks
31	Krogh 2009 Denmark	Outpatients Mean age: 39 (SD 9) 74% female	HAM-D <sub>17</sub> : 17.8 (SD 4)	165 (165)	<i>Exercise:</i> (1) Aerobic supervised group exercise OR (2) supervised group resistance training <i>Control group:</i> relaxation and stretching exercise.	(1)and (2): 2 sessions per week  Control group: 2 sessions per week	16 weeks
38	Mota-Pereira 2011 Portugal	Outpatients Treatment resistant Mean age: 47.5 (SD 3) 65.5% female	HAM-D <sub>17</sub> : 17.1 (SD 3)	33 (29)	<i>Aerobic exercise:</i> Homebased exercise + supervised. <i>Control group:</i> Attention control.	4 home-based sessions/week. 1 supervised session/week Control group: 1 supervised session/week	12 weeks
45	Krogh 2012 Denmark	Outpatients Mean age: 42 (SD 11) 67% female	HAM-D <sub>17</sub> : 18.9 (SD 4)	115 (115)	<i>Aerobic exercise:</i> Supervised group exercise. <i>Control group:</i> Supervised stretching exercise in groups.	3 sessions per week  Control group: 3 sessions per week	12 weeks
50	Chalder 2012 UK	Outpatients Mean age: 40 (SD 13) 66% female	BDI: 32.1 (SD 9)	361 (361)	<i>Exercise:</i> Participants received individually tailored support and encouragement to engage in physical activity. <i>Control group:</i> Standard GP care.	Individual	16 weeks

Fang 2013 China	Inpatients Mean age: 44 (SD 14) 66.9% female	HAM-D <sub>24</sub> : 29.2 (SD 5)	90 (90)	<i>Aerobic exercise:</i> Group 1 and 2 had supervised group exercise, high intensity. <i>Control group:</i> 15 min stretching	Group 1 and 2 had 3 and 5 sessions per week, respectively Control group: 3 sessions per week	6 weeks
Huipeng 2013 China	Inpatients Mean age: 30 (SD 5) 100% female	HAM-D <sub>17</sub> : 28 (SD 5)	68 (68)	<i>Aerobic exercise:</i> Jogging <i>Control group:</i> Standard treatment	5 sessions per week	6 weeks
Cassandra 2014 Honkong	Inpatients Mean age: 46 (SD 12) 67.3% female	MADRS: 19 (10)	52 (52)	<i>Aerobic exercise:</i> Supervised exercise. <i>Control group:</i> 10 min stretching.	5 sessions per week	3 weeks
Danielsson 2014 Sweden	Outpatients Mean age: 45 (SD 13) 76% female	MADRS: 24.0 (SD 5)	42 (42)	<i>Mixed aerobic and non-aerobic exercise:</i> First two weeks individual supervised exercise then supervised group exercise. <i>Control group:</i> One session with advice on physical activity.	2 sessions per week	10 weeks
Pfaff 2014 Australia	Outpatients Mean age: 61 (SD 8) 63% female	MADRS: 21.3 (SD NR)	200 (200)	<i>Resistance exercise:</i> Supervised home-based exercise <i>Control group:</i> Standard GP care	3 sessions per week	12 weeks
Guifeng 2015 China	Inpatients Mean age: 33 (SD 14) 70% female	HAM-D <sub>24</sub> : 25.9 (SD 4)	70 (70)	<i>Aerobic exercise:</i> Supervised group exercise <i>Control group:</i> Standard treatment	5 sessions per week	8 weeks
Junchin 2015 China	Inpatients Mean age: 28 (SD 7) 61% female	HAM-D <sub>24</sub> : 25.8 (SD 3)	70 (70)	<i>Aerobic exercise:</i> Supervised aerobic exercise of the patients own choice  <i>Control group:</i> Standard treatment	5 sessions per week	8 weeks
Schuch 2015 Brazil	Inpatients Mean age: 40 (SD 11) 74% female	HAM-D <sub>17</sub> : 26.7 (SD 2)	50 (50)	<i>Aerobic exercise:</i> Supervised individual exercise. <i>Control group:</i> Standard treatment.	3 sessions per week	2 weeks
Kerling 2015 Germany	Inpatients Mean age: 43 (SD 10)	MADRS: 24.0 (SD 9)	42 (42)	<i>Aerobic exercise:</i> Supervised exercise. <i>Control group:</i> Standard treatment.	3 sessions per week	6 weeks
Belvederi 2015 Italy	Outpatients Mean age: 75 (SD 6) 71% female	HAM-D <sub>17</sub> : 20.1 (SD 3)	121 (121)	<i>Aerobic exercise:</i> (1) Sertraline + supervised non-progressive exercise OR (2) sertraline + supervised progressive aerobic	3 sessions per week	24 weeks

					exercise. <i>Control group:</i> Sertraline.		
Carneiro 2015 Portugal	Outpatients Mean age: 50.16 (SD 12) 100% female	BDI: 48.8 (SD 10)	26 (19)	<i>Aerobic exercise:</i> Supervised exercise <i>Control group:</i> Standard treatment	3 sessions per week	16 weeks	
Doose 2015 Germany	Outpatients Mean age: 47.9 (SD 11) 63% female	HAM-D <sub>17</sub> : 14.2 (SD 3)	46 (46)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Standard treatment	3 sessions per week	8 weeks	
Pentecost 2015 UK	Outpatients Mean age: 44.4 (SD 14) 48% female	PHQ-9: 16.5 (SD 4)	60 (44)	<i>Exercise:</i> Behavioral activation plus physical activity promotion <i>Control group:</i> Behavioral activation	Individual	12 weeks	
Salehi 2016 Iran	Inpatients Mean age: 30.0 (SD 6) 35% female	HAM-D <sub>21</sub> : 43.4 (SD 8)	40 (40)	<i>Aerobic exercise + ECT:</i> Supervised aerobic exercise <i>Control group:</i> ECT	3 sessions per weeks Control group 3 ECTs per week	4 weeks	
Legrand 2016 France	Inpatients Mean age: 46.9 (SD 13) 67% female	BDI: 36.0 (SD 6)	24 (24)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Standard treatment	10 sessions in 10 consecutive days	10 days	
Euteneuer 2017 Germany	Outpatients Mean age: 37.1 (SD 12) 52% female	BDI: 27.2 (SD 9)	71 (68)	<i>Exercise:</i> CBT + PA promotion <i>Control group:</i> CBT + low energy activities	Individual	16 weeks	
Olson 2017 Ireland	Outpatients Mean age: 21.1 (SD 2) 80% female	BDI: 24.2 (SD 12)	50 (30)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Stretching exercise	3 sessions per week 3 sessions per week	8 weeks	
Patten 2017 USA	Outpatients Mean age: 37.5 (SD 11) 100% female	PHQ-9: 11.7 (SD 5)	30 (26)	<i>Aerobic exercise:</i> Supervised aerobic exercise <i>Control group:</i> Health education	3 sessions per week	12 weeks	

SCL-D: Symptom Check List, depression subscale; HAM-D<sub>17</sub>: Hamilton Depression Scale, 17 items; BDI: Beck's Depression Inventory; SD: Standard deviation; ECT: Electroconvulsive therapy; PHQ-9: Patient Health Questionnaire; CBT: Cognitive Behavioral Therapy

**Table 2.** Risk of bias in trials assessing exercise for patients diagnosed with depression

Author, Year of publication	Sequence generation	Allocation concealment	Blinding of participants and trial personnel assessors	Blinding of outcome assessors	Incomplete outcome data	Selective outcome reporting	For profit bias	Other bias	Comment on 'Other bias'
Klein 1985	Unclear	Unclear	High	High	High	Low	Low	Low	
Martinsen 1985	Unclear	Unclear	High	High	High	Low	High	Low	
Epstein 1986	Unclear	Unclear	High	High	High	Low	Unclear	High	Baseline difference
Doyne 1987	Unclear	Unclear	High	Low	High	Low	Unclear	High	Baseline difference
Veale 1992	Unclear	Unclear	High	High	High	Low	Low	High	Baseline difference
Singh 1997	Low	Unclear	High	Low	Low	Low	Low	High	Baseline difference
Blumenthal 1999	Unclear	Unclear	High	Low	High	Low	High	Low	
Mather 2002	Low	Low	High	Low	High	Low	Low	Low	
Dunn 2005	Low	Low	High	Low	High	High	High	Low	
Singh 2005	Low	Low	High	Low	High	Low	Unclear	Low	
Pilu 2007	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Viera 2007	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Blumenthal 2007	Low	Low	High	Low	High	High	Low	Low	
Krogh 2009	Low	Low	High	Low	Low <sup>1</sup>	High	High	High	Baseline difference
Mota-Pereira 2011	Unclear	Unclear	High	Low	High	Low	High	High	Baseline difference
Krogh 2012	Low	Low	High	Low	Low	Low	Low	Low	
Chalder 2012	Low	Low	High	High	Low	Low	Low	Low	
Fang 2013	Unclear	Unclear	High	Unclear	Unclear	High	Unclear	Low	
Huipeng 2013	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Cassandra 2014	Low	Unclear	High	Low	High	Low	Low	Low	
Danielsson 2014	Unclear	Low	High	Low	High	Low	Low	Low	
Pfaff 2014	Low	Low	High	Low	Low <sup>1</sup>	Low	Low	High	Baseline difference
Guifeng 2015	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Jinchun 2015	Unclear	Unclear	High	Unclear	Low	Low	Unclear	Low	
Schuch 2015	Unclear	Low	High	Low	Low	Low	Low	Low	
Kerling 2015	Unclear	Unclear	High	Unclear	Low	Low	Low	Low	
Belvederi 2015	Low	Low	High	Low	High	Low	Low	High	Post-hoc sample size
Carneiro 2015	Unclear	Low	High	High	Unclear	Low	Low	Low	
Doose 2015	Unclear	Unclear	High	High	High	Low	Low	High	No sample size calc.
Pentecost 2015	Low	Low	High	High	High	Low	Low	Low	
Salehi 2016	High	High	High	Low	Unclear	Low	Low	High	Baseline difference
Legrand 2016	Low	High	High	High	High	Low	Unclear	Low	
Euteneuer 2017	Low	Unclear	High	High	High	Low	Low	Low	
Olson 2017	Low	Unclear	High	High	High	Low	Low	Low	
Patten 2017	Unclear	Unclear	High	High	High	Low	Low	Low	



**Table 3.** Heterogeneity of effect estimates for trials assessing the effect of exercise for patients diagnosed with depression explored by comparing sub-groups.

Subgroups	Number of Trials (participants)	Random effects meta-analysis SMD (95% CI., p, I <sup>2</sup> )	Subgroup explains heterogeneity P value
<b>Risk of bias</b>			
Less than high risk of bias <sup>1</sup>	4 (530)	-0.11 (-0.41 to 0.18; p = 0.45; I <sup>2</sup> = 62%)	<0.001
High risk of bias	31 (1968)	-0.75 (-0.98 to -0.52; p < 0.001; I <sup>2</sup> = 81%)	
<b>Age</b>			
Old (>59 years)	5 (492)	-0.77 (-1.34 to -0.19; p = 0.009; I <sup>2</sup> = 87%)	0.78
Young (<59 years)	30 (2006)	-0.68 (-0.90 to -0.45; p < 0.001; I <sup>2</sup> = 83%)	
<b>Exercise context</b>			
Group exercise	26 (1785)	-0.75 (-1.01 to -0.50; p < 0.001; I <sup>2</sup> = 83%)	0.30
Individual exercise	9 (713)	-0.52 (-0.88 to -0.16; p = 0.005; I <sup>2</sup> = 73%)	
<b>Duration</b>			
Less than 10 weeks	15 (721)	-0.92 (-1.09 to -0.74; p < 0.001; I <sup>2</sup> = 14%)	0.007
10 weeks or more	20 (1777)	-0.49 (-0.75 to -0.23; p < 0.001; I <sup>2</sup> = 83%)	
<b>Attention control</b>			
Attention control	10 (733)	-0.56 (-0.98 to -0.15; p = 0.008; I <sup>2</sup> = 85%)	0.91
Waitlist	2 (47)	-0.67 (-2.48 to 1.13; p = 0.47; I <sup>2</sup> = 88%)	
<b>Pharmacotherapy</b>			
Add-on	11 (734)	-0.92 (-1.38 to -0.46; p < 0.001; I <sup>2</sup> = 86%)	0.82
No medication	6 (318)	-0.82 (-1.58 to -0.06; p = 0.03; I <sup>2</sup> = 88%)	
<b>Somatic comorbidity</b>			
Somatic co-morbidity	0	N/A	
No co-morbidity	35 (2331)	N/A	
<b>Minor depression</b>			
Incl. minor depression	6 (350)	-0.90 (-1.65 to -0.15; p = 0.02; I <sup>2</sup> = 86%)	0.53
No minor depression	25 (2148)	-0.65 (-0.87 to -0.43; p < 0.001; I <sup>2</sup> = 81%)	
<b>Patient setting</b>			
Inpatients	10 (549)	-0.88 (-1.07 to -0.70; p < 0.001; I <sup>2</sup> = 6%)	0.07
Outpatients	21 (1782)	-0.60 (-0.85 to -0.35; p < 0.001; I <sup>2</sup> = 83%)	
<b>Trial size</b>			
Trials n ≤ 50	18 (578)	-1.11 (-1.52 to -0.72; p < 0.001; I <sup>2</sup> = 78%)	0.001
Trials n > 50	17 (1920)	-0.37 (-0.57 to -0.18; p < 0.001; I <sup>2</sup> = 75%)	
<b>Increase in exercise capacity</b>			
VO <sub>2</sub> max > 2.8 ml/kg/min	5 (340)	-0.48 (-1.08 to 0.13; p = 0.12; I <sup>2</sup> = 86%)	0.65
VO <sub>2</sub> max ≤ 2.8 ml/kg/min	6 (661)	-0.32 (-0.61 to 0.02; p = 0.03; I <sup>2</sup> = 68%)	

**Table 4.** Summary of findings

<b>Exercise compared to control or treatment as usual for depression</b>						
<b>Patient or population:</b> depression						
<b>Setting:</b> In- or out-patients						
<b>Intervention:</b> exercise						
<b>Comparison:</b> control or treatment as usual						
Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Risk with control or treatment as usual	Risk with exercise				
Severity of depression	-	<b>0.66 SMD lower</b> (0.46 lower to 0.86 lower)	-	2498 (35 RCTs)	⊕○○○ VERY LOW <sup>1</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.
Lack of remission	<b>Study population</b>		<b>RR 0.78</b> (0.68 to 0.90)	1639 (19 RCTs)	⊕○○○ VERY LOW <sup>2</sup>	Remission is, with minor variations, defined as not full-filling the criteria for depression.
	646 per 1000	<b>504 per 1000</b> (426 to 594)				
Serious adverse events	<b>Study population</b>		<b>RR 2.21</b> (0.24 to 20.21)	335 (3 RCTs)	⊕⊕○○ LOW <sup>3</sup>	
	0 per 1000	<b>0 per 1000</b> (0 to 0)				
Quality of life	-	<b>0.40 SMD higher</b> (0.03 lower to 0.83 higher)	-	827 (9 RCTs)	⊕○○○ VERY LOW <sup>4</sup>	Quality of life was assessed using a number of different methods. Higher score indicates improved quality of life. Seven of 24 trials reported on this outcome
Depression severity after the intervention	-	<b>0.06 SMD lower</b> (0.25 lower to 0.14 higher)	-	713 (7 RCTs)	⊕⊕○○ LOW <sup>5</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.
Lack of remission after the intervention	<b>Study population</b>		<b>RR 0.95</b> (0.82 to 1.11)	777 (5 RCTs)	⊕⊕○○ LOW <sup>6</sup>	
	469 per 1000	<b>446 per 1000</b> (385 to 521)				
Depression severity. Restricted to trials with less than high risk of bias.	-	<b>0.11 SMD lower</b> (0.41 lower to 0.18 higher)	-	530 (4 RCTs)	⊕⊕○○ LOW <sup>7</sup>	Lower depression scores indicate improvement. SMD of 0.3 is considered clinically relevant.

\*The risk in the intervention group (and its 95% confidence interval) is based on the assumed risk in the comparison group and the relative effect of the intervention (and its 95% CI). CI: Confidence interval; SMD: Standardised mean difference; RR: Risk ratio

**GRADE Working Group grades on evidence**

<b>High quality:</b> We are very confident that the true effect lies close to that of the estimate of the effect	<b>High</b>
We are moderately confident in the effect estimate. The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different	<b>Moderate quality:</b>
the effect estimate is limited: The true effect may be substantially different from the estimate of the effect	<b>Low quality:</b> Our confidence in
in the effect estimate: The true effect is likely to be substantially different from the estimate of the effect	<b>Very low:</b> We have very little confidence

1. Downgraded by 3: risk of bias, inconsistency and publication bias
2. Downgraded by 3: risk of bias, inconsistency and publication bias
3. Downgraded by 2: imprecision and publication bias

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

- 4. Downgraded by 3: risk of bias, inconsistency and imprecision
- 5. Downgraded by 2: risk of bias and imprecision
- 6. Downgraded by 2: risk of bias and imprecision
- 7. Downgraded by 2: inconsistency and imprecision

For peer review only

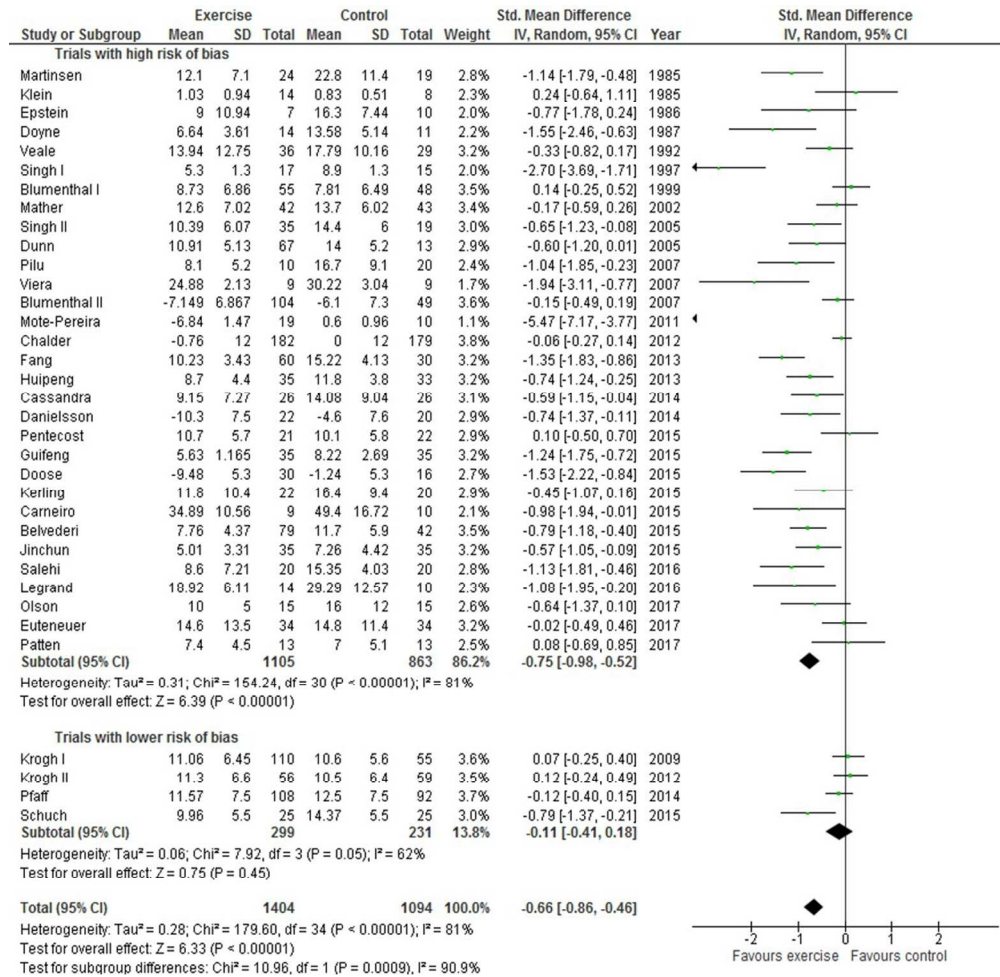


Figure 1

94x91mm (300 x 300 DPI)

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Table S1**

**Table S1.** Imputation of missing data for trials assessing exercise for patients diagnosed with depression

Outcome	Result from review	Best/worse (1SD)	Best/worse (2SD)	Worse/best (1SD)	Worse/best (2SD)
Depression SMD (95% CI)	-0.66 (-0.86 to -0.45) p < 0.001; I <sup>2</sup> = 81%	-0.77 (-1.00 to -0.54) p < 0.001; I <sup>2</sup> = 86%	-0.78 (-1.02 to -0.55) p < 0.001; I <sup>2</sup> = 86%	-0.60 (-0.81 to -0.39) p < 0.001; I <sup>2</sup> = 84%	-0.57 (-0.78 to -0.36) p < 0.001; I <sup>2</sup> = 84%
		<b>Good Outcome</b>	<b>Poor outcome</b>	<b>Good/poor outcome</b>	<b>Poor/good outcome</b>
Lack of remission (95% CL)	RR 0.78 (0.68 to 0.90) p < 0.001; I <sup>2</sup> = 69%	RR 0.75 (0.64 to 0.89) p = 0.0008; I <sup>2</sup> = 73%	RR 0.88 (0.83 to 0.94) p = 0.0002; I <sup>2</sup> = 69%	RR 0.71 (0.61 to 0.81) p < 0.001; I <sup>2</sup> = 68%	RR 0.86 (0.71 to 1.04) p = 0.12; I <sup>2</sup> = 83%
Serious adverse events (95% CL)	RR 2.21 (0.24 to 20.21) p = 0.48; I <sup>2</sup> = 0%	RR 2.19 (0.23 to 20.76) p = 0.50; I <sup>2</sup> = 50%	RR 0.92 (0.37 to 2.30) p = 0.86; I <sup>2</sup> = 60%	RR 0.08 (0.02 to 0.34) p = 0.001; I <sup>2</sup> = 5.4%	RR 19.17 (2.64 to 139.2) p = 0.004; I <sup>2</sup> = 0%

Imputation of missing data for continuous outcome: 'best-worst' - assumed that all participants lost to follow-up in the intervention group had a beneficial outcome (the group mean minus 1 or 2 SD), and all participants lost to follow-up in the placebo group have had a harmful outcome (the group mean plus 1 SD and 2 SD). The reverse 'worst-best-case' scenario is the reverse of the 'best-worst' scenario.

Missing data for the 'remission' outcome was imputed according to the following scenarios: 1) poor outcome analysis: none of the drop-outs/participants lost from both arms experienced the outcome; 2) good outcome analysis: all of the drop-outs/participants lost from both arms experienced the outcome; 3) extreme case analysis favouring the experimental intervention, all of the drop-outs/participants lost from the experimental arm, but none of the drop-outs/participants lost from the control arm experienced the outcome; and 4) extreme case analysis favouring the control: all drop-outs/participants lost from the experimental arm, but none from the control arm experienced the outcome. Missing data for 'serious adverse events' was calculated with the reverse assumptions.

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Table S2**

**Table S2.** Trials characteristics for exploration of heterogeneity in trials assessing the effect of exercise in patients diagnosed with depression

Trial	Lower risk of bias	Age > 60	Group vs. individual	Duration	Attention control waitlist	Exercise as add on to drugs vs. exercise alone	Within-study dose exercise	Increase in VO2max <sup>1</sup>	Somatic disease vs. only MD	Trial Includes minor depression
Klein 1985	No	Young	Individual	12 weeks	Other	Exercise alone	No	No	No	Yes
Martinsen 1985	No	Young	Group	9 weeks	Attention control	Unclear	No	11 <sup>a</sup>	No	No
Epstein 1986	No	Young	Group	8 weeks	Waitlist	Unclear	No	No	No	Yes
Doyne 1987	No	Young	Individual	8 weeks	Waitlist	Exercise alone	No	No	No	Yes
Veale 1992	No	Young	Group	12 weeks	Other	Unclear	No	No	No	No
Singh 1997	No	Old	Group	10 weeks	Attention control	Exercise alone	No	N/A	No	Yes
Blumenthal 1999	No	Young	Group	16 weeks	Other	Add on	No	2.3	No	No
Mather 2002	No	Old	Group	10 weeks	Attention control	Add on	No	No	No	No
Dunn 2005	No	Young	Individual	12 weeks	Attention control	Exercise alone	Yes	No	No	No
Singh 2005	No	Old	Group	8 weeks	Other	Exercise alone	Yes	N/A	No	Yes
Pilu 2007	No	Young	Group	24 weeks	Other	Add on	No	No	No	No
Viera 2007	No	Young	Group	12 weeks	Other	Add on	No	No	No	No
Blumenthal 2007	No	Young	Group	16 weeks	Other	Add on	No	2.0 <sup>a</sup>	No	No
Krogh 2009	Yes	Young	Group	16 weeks	Attention control	No	No	2.9	No	No
Mota-Pereira 2011	No	Young	Group	12 weeks	Other	Add on	No	No	No	No
Krogh 2012	Yes	Young	Group	12 weeks	Attention control	Exercise alone	No	3.4	No	No
Chalder 2012	No	Young	Individual	32 weeks	Other	No	No	No	No	No
Fang 2013	No	Young	Group	6 weeks	Attention control	No	Yes	No	No	No
Huipeng 2013	No	Young	Group	6 weeks	Other	No	No	No	No	No
Cassandra 2014	No	Young	Group	3 weeks	Other	Add on	No	No	No	No
Danielsson 2014	No	Young	Group	10 weeks	Other	Add on	No	2.4	No	No
Pfaff 2014	Yes	Old	Group	12 weeks	Other	No	No	1.5	No	Yes
Guifeng 2015	No	Young	Group	8 weeks	Other	No	No	No	No	No
Jinchun 2015	No	Young	Group	8 weeks	Other	No	No	No	No	No
Schuch 2015	Yes	Young	Individual	2 weeks	Other	No	No	No	No	No
Kerling 2015	No	Young	Group	6 weeks	Other	No	No	2.8	No	No

1	Belvederi	No	Old	Group	24 weeks	Other	Add on	Yes	0.3 <sup>a</sup>	No	No
2	2015										
3	Carneiro	No	Young	Group	16 weeks	Other	Add on	No	No	No	No
4	2015										
5	Doose	No	Young	Group	8 weeks	Other	No	No	3.2	No	No
6	2015										
7	Pentecost	No	Young	Individual	12 weeks	Other	No	No	No	No	No
8	2015										
9	Legrand	No	Young	Individual	10 days	Other	No	No	No	No	No
10	2016										
11	Salehi	No	Young	Individual	4 weeks	Other	Add on	No	No	No	No
12	2016										
13	Euteneuer	No	Young	Individual	16 weeks	Attention	No	No	No	No	No
14	2017					control					
15	Olsen	No	Young	Group	8 weeks	Attention	No	No	No	No	No
16	2017					control					
17	Patten	No	Young	Group	12 weeks	Other	No	No	5.0	No	No
18	2017										

<sup>1</sup>Increase in VO<sub>2</sub>max is based on increase in intervention group, if <sup>a</sup> then value is based on an estimate from text or figures.

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplementary Table

**Table S3.** Heterogeneity of effect estimates for trials assessing the effect of exercise for patients diagnosed with depression on lack of remission.

Subgroups	Number of Trials (participants)	Random effects meta-analysis RR (95% CI., p, I <sup>2</sup> )	Subgroup explains heterogeneity P value
<b>Risk of bias</b>			
Less than high risk of bias <sup>1,2</sup>	2 (165)	0.95 (0.74 to 1.23; p = 0.70; I <sup>2</sup> = 20%)	0.18
High risk of bias	17 (1474)	0.77 (0.64 to 0.92; p = 0.003; I <sup>2</sup> = 75%)	
<b>Age</b>			
Old (>59 years)	3 (299)	0.61 (0.21 to 1.02; p = 0.37; I <sup>2</sup> = 91%)	0.62
Young (<59 years)	16 (1340)	0.81 (0.70 to 0.93; p = 0.003; I <sup>2</sup> = 64%)	
<b>Exercise context</b>			
Group exercise	14 (1156)	0.80 (0.66 to 0.96; p = 0.02; I <sup>2</sup> = 72%)	0.69
Individual exercise	5 (483)	0.74 (0.52 to 1.04; p = 0.08; I <sup>2</sup> = 77%)	
<b>Duration</b>			
Less than 10 weeks	8 (393)	0.63 (0.51 to 0.77; p < 0.001; I <sup>2</sup> = 40%)	0.004
10 weeks or more	11 (1246)	0.93 (0.78 to 1.10; p = 0.39; I <sup>2</sup> = 69%)	
<b>Attention control</b>			
Attention control	4 (364)	0.91 (0.73 to 1.12; p = 0.38; I <sup>2</sup> = 42%)	0.07
Waitlist	1 (25)	0.44 (0.21 to 0.93; p = 0.03; I <sup>2</sup> = 0%)	
<b>Pharmacotherapy</b>			
Add-on	7 (540)	0.72 (0.54 to 0.96; p = 0.03; I <sup>2</sup> = 69%)	0.62
No medication	4 (252)	0.75 (0.52 to 1.09; p = 0.13; I <sup>2</sup> = 66%)	
<b>Somatic comorbidity</b>			
Somatic co-morbidity	0	N/A	
No co-morbidity	19 (1639)	N/A	
<b>Minor depression</b>			
Incl. minor depression	3 (203)	0.63 (0.21 to 1.89; p = 0.41; I <sup>2</sup> = 87%)	0.69
No minor depression	16 (1436)	0.79 (0.68 to 0.92; p = 0.002; I <sup>2</sup> = 69%)	
<b>Patient setting</b>			
Inpatients	6 (322)	0.71 (0.60 to 0.84; p < 0.001; I <sup>2</sup> = 0%)	0.21
Outpatients	13 (1317)	0.84 (0.69 to 1.01; p = 0.07; I <sup>2</sup> = 77%)	
<b>Trial size</b>			
Trials n ≤ 52	9 (358)	0.62 (0.50 to 0.76; p < 0.001; I <sup>2</sup> = 45%)	0.002
Trials n > 52	10 (1281)	0.95 (0.80 to 1.12; p = 0.52; I <sup>2</sup> = 68%)	

<sup>1</sup>Trials potentially having less bias than trials with high risk of bias.

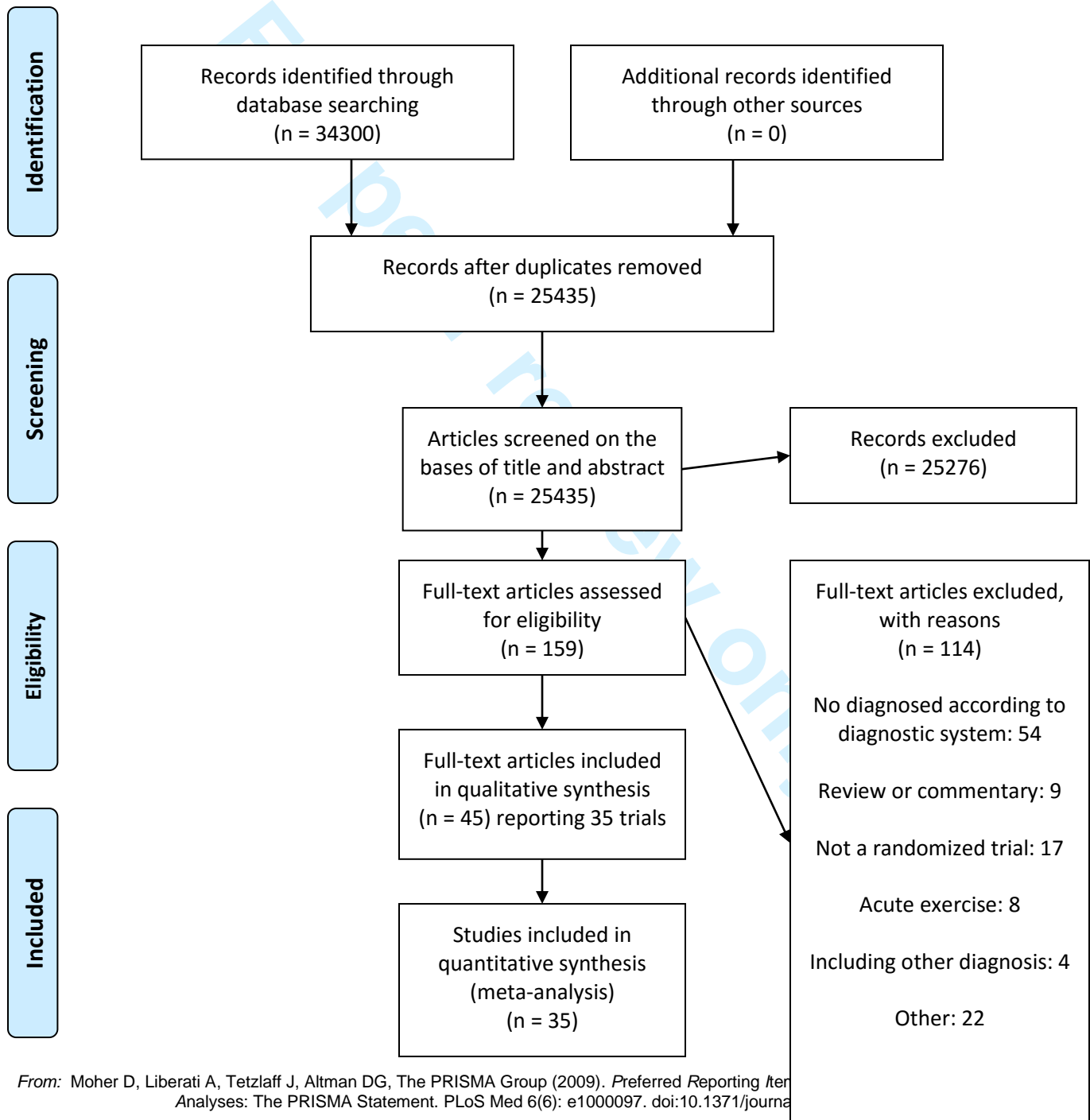


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

### Supplemental

Figure S1. Flow diagram for identification of trials assessing the effects of exercise for patients with depression.



For more information, visit [www.prisma-statement.org](http://www.prisma-statement.org).

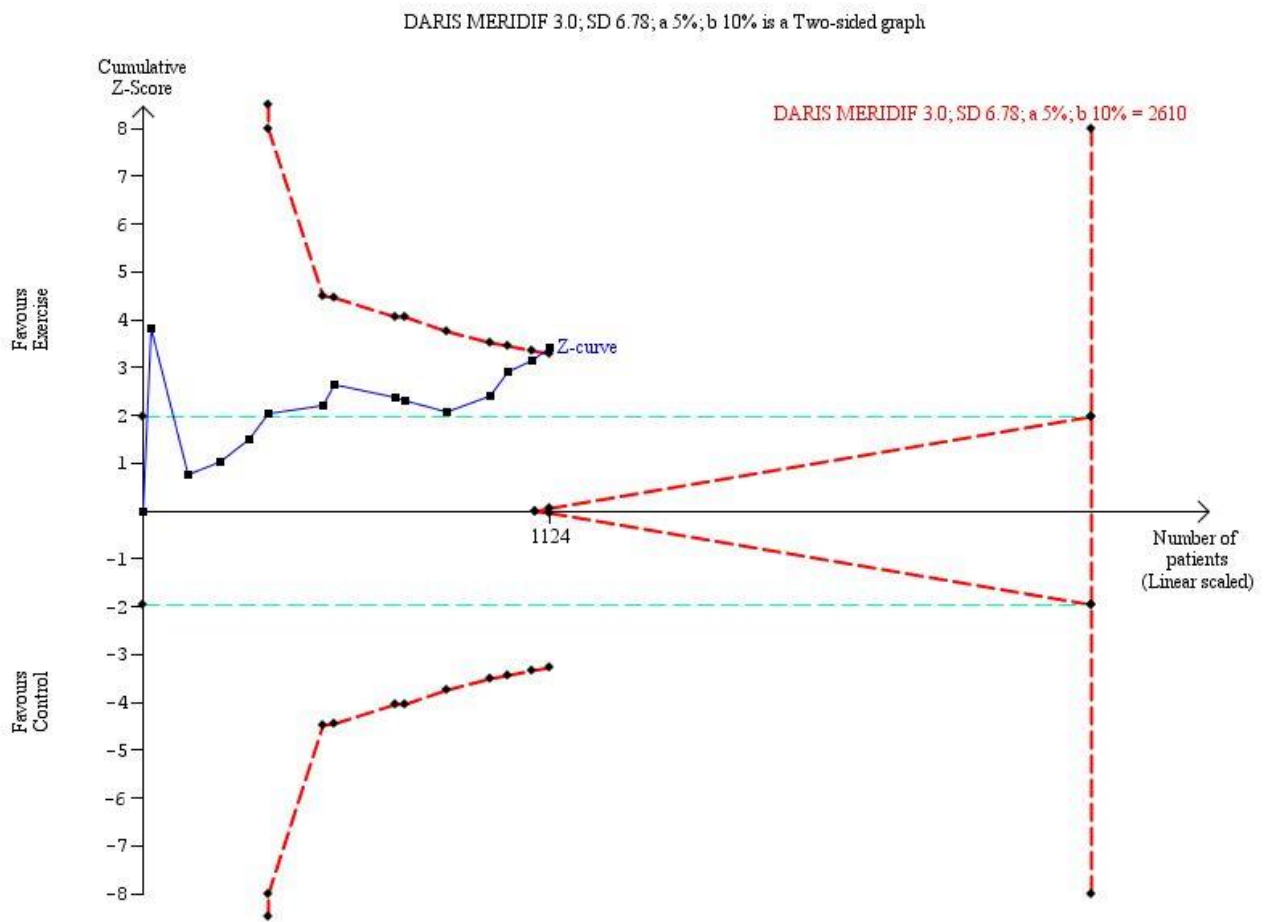
For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S2. Trial Sequential Analysis and required information size for the effect of exercise for depressive symptoms including four-teen trials reporting on HAM-D<sub>17</sub>.

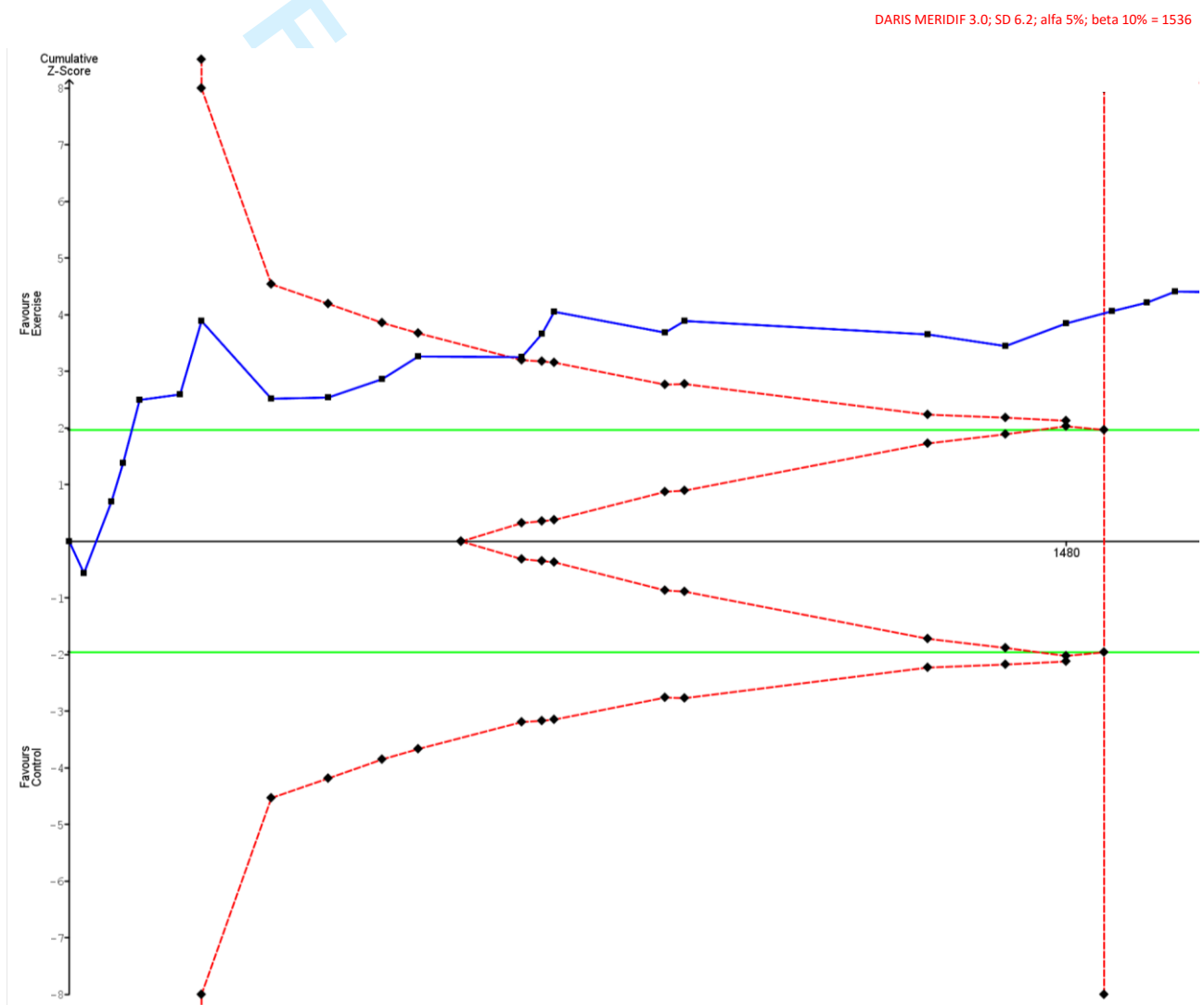


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S3. Trial Sequential Analysis and required information size for the effect of exercise for depressive symptoms including 35 trials 'converted' to a HAM-D<sub>17</sub> scale.



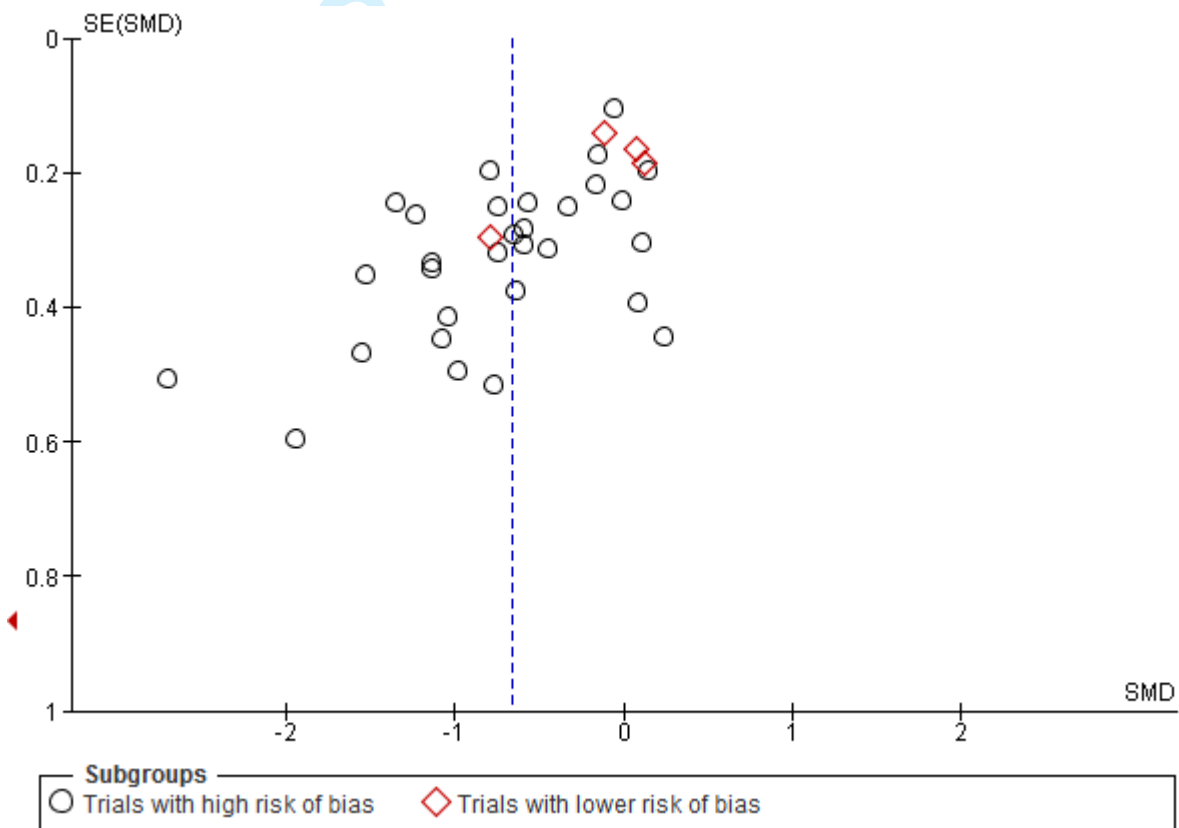
Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Figure**

**Figure S4.**

Funnel plot of 35 trials assessing the antidepressant effect of exercise as a continuous outcome

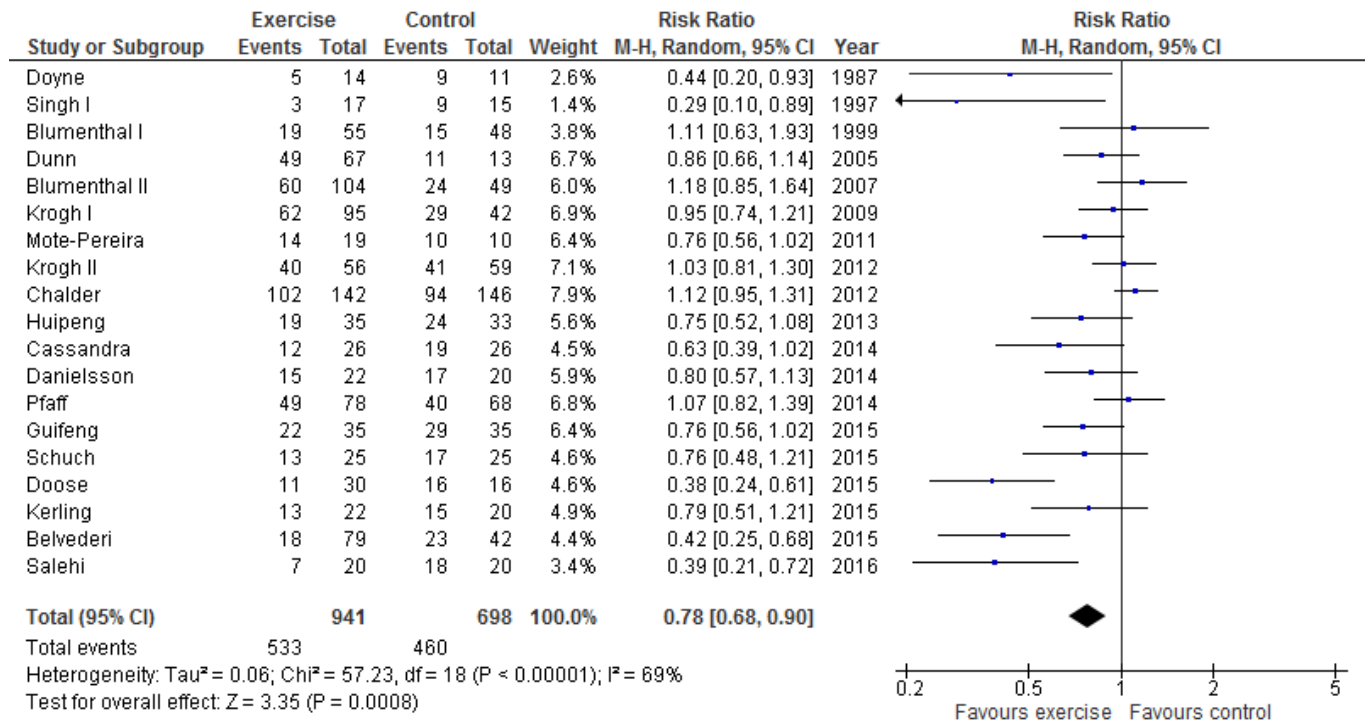


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S5. Effect of exercise on lack of remission for patients diagnosed with depression



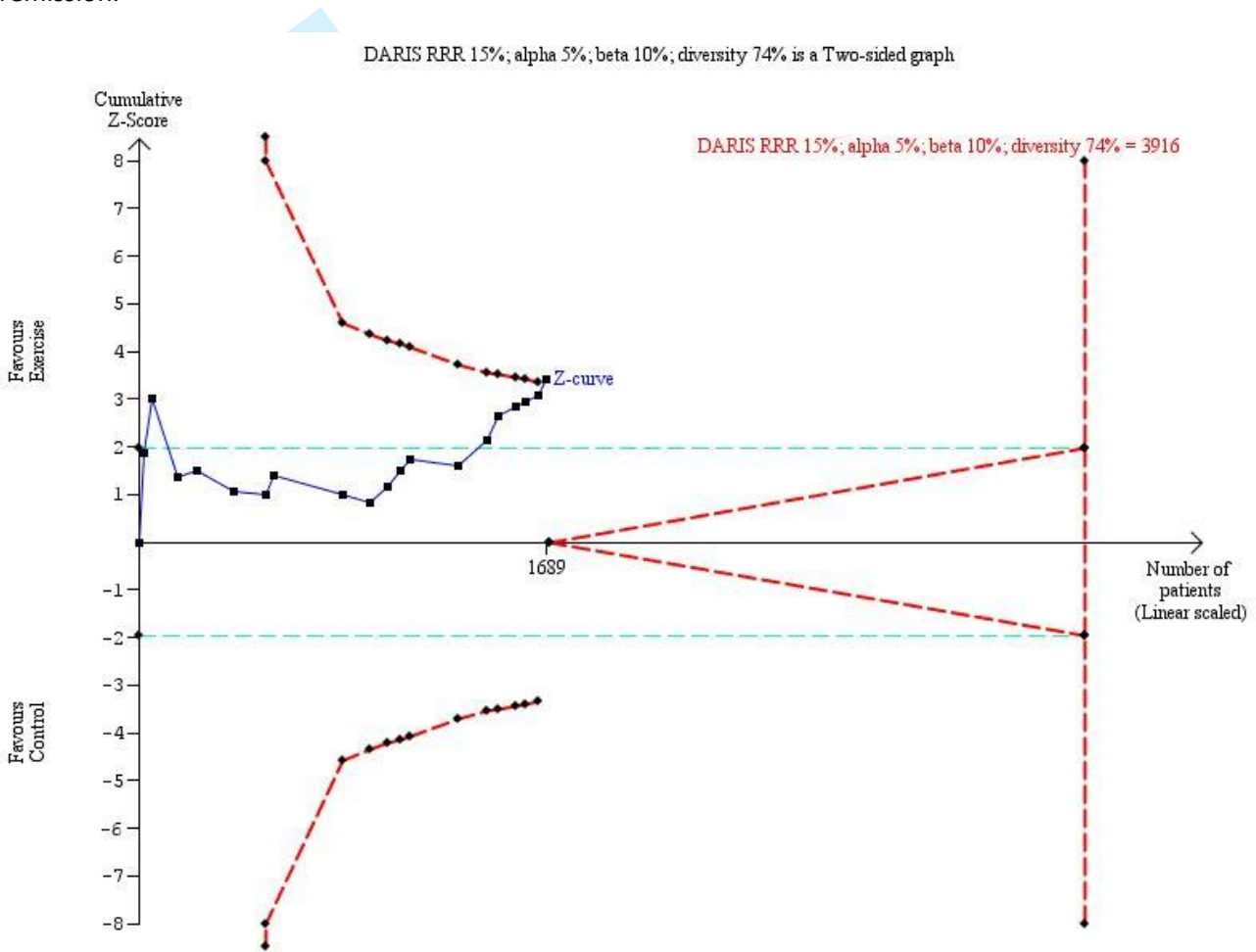
view only

Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

Supplementary Figure

Figure S6. Trial Sequential Analysis and required information size for the effect of exercise on lack of remission.

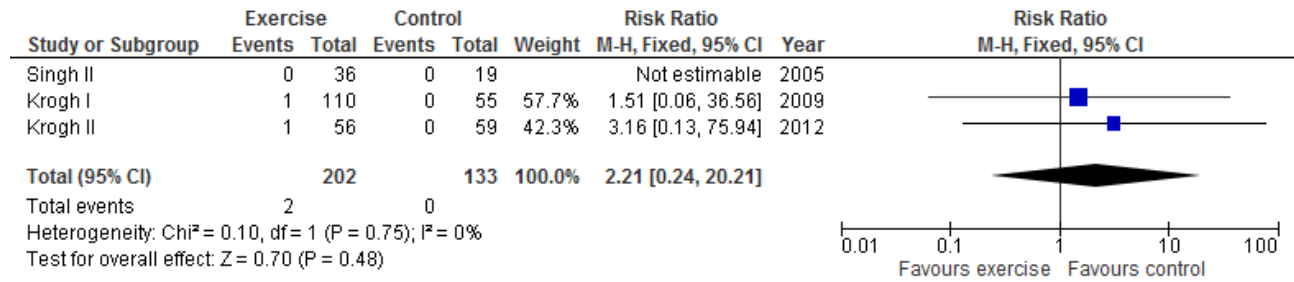


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

**Supplementary Figure S7**

**Figure S7.** Effect of exercise on risk of serious adverse events for patients diagnosed with depression

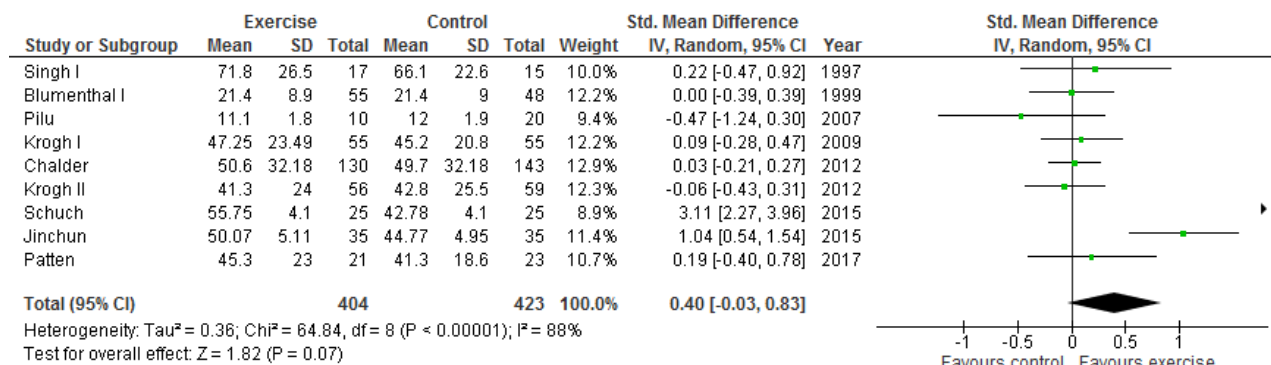


Article:

Exercise for patients with major depression: A systematic review with meta-analysis and trial sequential-analysis

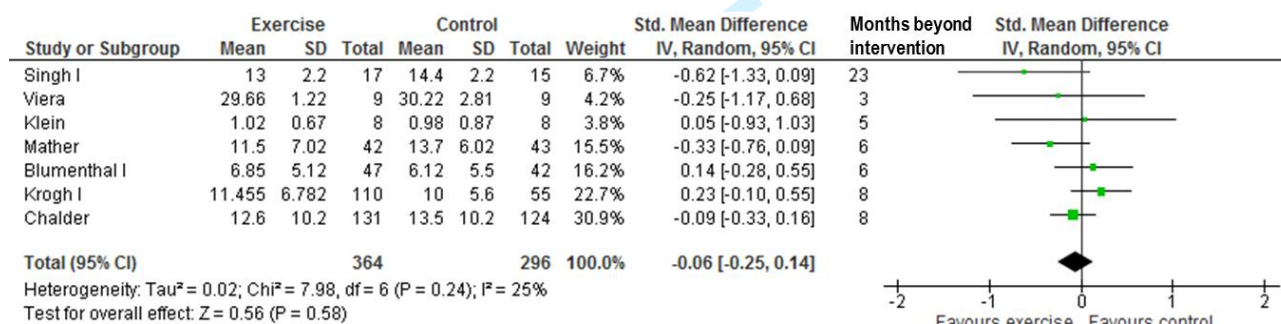
Supplementary Figure S8-S10

Figure S8. The effect of exercise on quality of life in patients diagnosed with depression



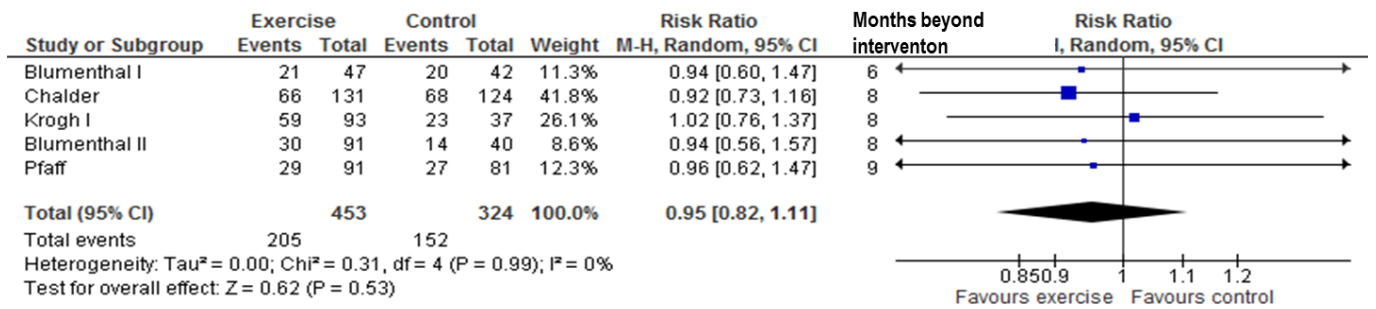
Quality of life was assessed using different scales: Singh I, Chalder and Patten used the SF-36, Blumenthal used Life Satisfaction Index, Pilu and Schuch used the WHOQOL, Krogh I and Krogh II used the WHO-Five Well-being Scale, and Jinchun used the GQOLI-74.

Figure S9. The effect of exercise on depression severity after the intervention in patients diagnosed with depression





**Figure S10.** The effect of exercise on risk of lack of remission after the intervention in patients diagnosed with depression



For peer review only

1  
2  
3  
4 Article:

5  
6 Exercise for patients with major depression: A systematic review with meta-analysis and trial  
7 sequential-analysis  
8

9  
10 **Supplementary material (S1)**

11 An example of bibliographical search for PubMed

12  
13 #1 Depression [MeSH]

14  
15 #2 Depressive disorder [MeSH]

16  
17 #3 Exercise [Text Word]

18  
19 #4 Aerobic [Text Word]

20  
21 #5 Non-aerobic [Text Word]

22  
23 #6 Physical activity [Text Word]

24  
25 #7 Physical fitness [Text Word]

26  
27 #8 Walking [MeSH]

28  
29 #9 Jogging [MeSH]

30  
31 #10 Running [MeSH]

32  
33 #11 Bicycling [MeSH]

34  
35 #12 Swimming [MeSH]

36  
37 #13 Strength [Text Word]

38  
39 #14 Resistance [Text Word]

40  
41 #15 #1 OR #2

42  
43 #16 #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14

44  
45 #17 #15 AND #16  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



# PRISMA 2009 Checklist

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

Section/topic	#	Checklist item	Reported on page #
<b>TITLE</b>			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
<b>ABSTRACT</b>			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of what is already known.	5
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	6
<b>METHODS</b>			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	1
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	7
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	8
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	7
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	8
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	9
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., $I^2$ for each meta-analysis).	9



# PRISMA 2009 Checklist

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	8
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	11
<b>RESULTS</b>			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Fig. 1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	Table 1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	Table 2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	Fig 2
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	Fig 3-fig8
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	Table 2
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	Table 3
<b>DISCUSSION</b>			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	21
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	21
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	22
<b>FUNDING</b>			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	26

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: [www.prisma-statement.org](http://www.prisma-statement.org).

Page 2 of 2