

## Review article

# Effect of exercise on depression severity in older people: systematic review and meta-analysis of randomised controlled trials

Christopher Bridle, Kathleen Spanjers, Shilpa Patel, Nicola M. Atherton and Sarah E. Lamb

**Background**

The prevalence of depression in older people is high, treatment is inadequate, it creates a substantial burden and is a public health priority for which exercise has been proposed as a therapeutic strategy.

**Aims**

To estimate the effect of exercise on depressive symptoms among older people, and assess whether treatment effect varies depending on the depression criteria used to determine participant eligibility.

**Method**

Systematic review and meta-analysis of randomised controlled trials of exercise for depression in older people.

**Results**

Nine trials met the inclusion criteria and seven were

meta-analysed. Exercise was associated with significantly lower depression severity (standardised mean difference (SMD) =  $-0.34$ , 95% CI  $-0.52$  to  $-0.17$ ), irrespective of whether participant eligibility was determined by clinical diagnosis (SMD =  $-0.38$ , 95% CI  $-0.67$  to  $-0.10$ ) or symptom checklist (SMD =  $-0.34$ , 95% CI  $-0.62$  to  $-0.06$ ). Results remained significant in sensitivity analyses.

**Conclusions**

Our findings suggest that, for older people who present with clinically meaningful symptoms of depression, prescribing structured exercise tailored to individual ability will reduce depression severity.

**Declaration of interest**

None.

Depression is the most common mental illness among older people, and is associated with increased morbidity, premature mortality and greater healthcare utilisation.<sup>1–3</sup> Treatment of depression is inadequate for most older people, being complicated by poor recognition and an increased prevalence of medication side-effects, polypharmacy and poor adherence to treatment.<sup>4–6</sup> Depression is predicted to become the leading cause of disease burden among older people by 2020,<sup>7</sup> at which time one in five of the population will be aged over 60 years.<sup>8</sup> Effective treatment of depression in older people is a salient public health priority, for which exercise has been increasingly evaluated as a potential therapeutic strategy.<sup>9,10</sup> Findings from recent reviews, however, are difficult to interpret clinically, since they reflect qualitative syntheses of evidence from randomised and non-randomised trials, and trials in which pre-existing depression has not been an eligibility criterion.<sup>11,12</sup> There is uncertainty concerning the effect of exercise on depression among older people with clinically significant symptoms of depression. The aim of this study was to provide a clinically meaningful synthesis of evidence to support treatment decisions. The primary objective was to estimate the effect of exercise on depression severity among older people with clinically significant symptoms of depression. The secondary objective was to investigate any potential variation in treatment effect among pre-specified subgroups of the study stratified by depression eligibility criteria, specifically the selection of participants according to clinician-diagnosed depression or a symptom checklist threshold.

**Method****Eligibility criteria**

Studies were considered for inclusion if they were randomised controlled trials (RCTs) of exercise interventions for depression among older people. A trial was accepted as a RCT if the allocation

of participants to treatment and comparison groups was reported as randomised. Studies were considered for inclusion if the sample mean age was  $\geq 60$  years. Setting the minimum age criterion at  $\geq 60$  years is consistent with previous reviews<sup>11,12</sup> and the World Health Organization's classification of older age,<sup>13</sup> and linking age to the sample, rather than the individual, recognises that trials vary in the use and precise specification of the minimum age criterion. The review included studies in which participant eligibility required pre-existing depression determined by a clinically valid method of assessment, such as a clinical interview, clinician diagnosis or symptom checklist threshold. Trials of any exercise intervention compared with any concurrent control were eligible. Exercise was defined as any planned or structured movement of the body performed systematically in terms of frequency, intensity and duration. Included trials reported depression as an outcome assessed at follow-up of  $\geq 3$  months.

**Study identification**

To identify relevant published, unpublished and ongoing trials, as well as existing systematic reviews, the following electronic databases were searched from inception to January 2011: CDSR, DARE, UK-NRR, CCT, HSRProj, CENTRAL, Medline; Embase, PsycINFO, SSCI, SportsDiscus, AMED, CINAHL, BioMed Central, HealthPromis, Index of Conference Proceedings, Theses, SIGLE and GreyLit. Search parameters were adapted to database requirements, and combined exploded MeSH terms and text words related to exercise, depression and age (see online supplement). The bibliographies of all included studies and review articles were screened for further references. Search results were recorded to bibliographic software, and two reviewers independently screened each citation for potential relevance against eligibility criteria. For all potentially relevant citations, full-text papers were obtained and assessed against eligibility

criteria by two reviewers independently, with disagreements resolved by discussion.

### Data abstraction

Data were extracted by one reviewer and checked for accuracy by another, using a template that included: (a) design, for example depression eligibility criteria, sample size and recruitment context; (b) participants, for example age, gender and baseline depression; (c) intervention, for example type, frequency and format of exercise; (d) outcome, for example depression measure, follow-up schedule and depression severity (mean and standard deviation) for each group at each follow-up; and (e) process, for example number of eligible patients invited and, for the exercise group, adherence, including the criteria used and the level achieved.

Two reviewers independently assessed risk of bias in each trial according to the adequacy of sequence generation, allocation concealment, masking of outcome assessors, completeness of follow-up and analysis by intention to treat. Each component was assessed as either adequate, inadequate or unclear, using Cochrane risk of bias criteria.<sup>14</sup> Risk of bias in each study was classified as either low (all criteria graded adequate), moderate (one criterion graded inadequate, or two graded unclear) or high (two or more criteria graded inadequate, or more than two graded unclear).

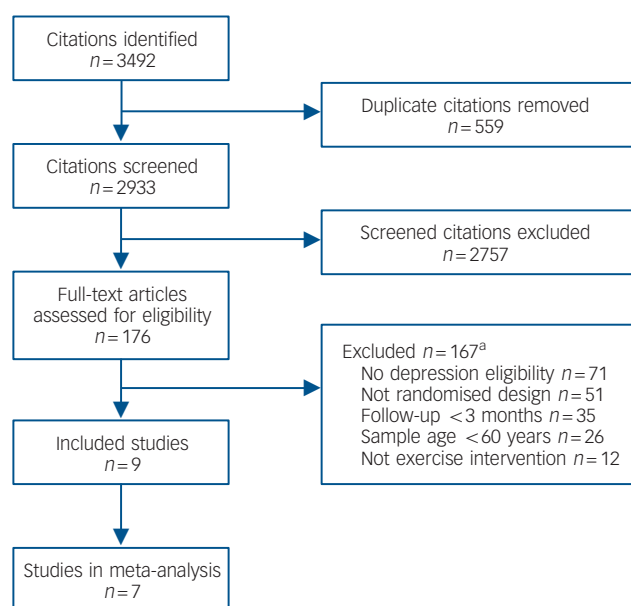
### Data analysis

All analyses were conducted using Review Manager version 5.1 software for Windows. All trials reported depression as a continuous outcome, but different measurers were used in the assessment. Thus, the summary measure of treatment effect was the between-groups difference in mean severity of depression, expressed as a standardised mean difference (SMD) using Hedges' (adjusted) *g*, which includes a correction term for sample size bias.<sup>15</sup> Statistical heterogeneity was assessed by the  $I^2$  test, which describes the percentage of variability among effect estimates beyond that expected by chance. Heterogeneity can be considered as unlikely to be important for  $I^2$  values up to 40%.<sup>14</sup> In the absence of statistical heterogeneity ( $I^2 \leq 40\%$ ), individual effect sizes were combined statistically using the inverse variance random-effects method, which assumes that true effects are normally distributed. The random-effects model is more conservative than the fixed-effect model since, by incorporating both within- and between-study variance, confidence intervals for the summary effect are wider. Risk of small study bias was assessed by visual assessment of funnel symmetry in the plots of each trial's SMD against its standard error (s.e.).<sup>14</sup>

The effect of exercise on depression severity was estimated in pre-specified subgroups of the study stratified by depression eligibility criteria. Specifically, we distinguished between trials in which participant eligibility was dependent on either satisfying clinical diagnostic criteria for depression or achieving a threshold score on a depression symptom checklist. The robustness of results was assessed in separate sensitivity analyses that excluded trials with moderate or high risk of bias, non-active or no intervention control comparators and end-points within rather than beyond the intervention period.

## Results

After removal of duplicates, the search strategy identified 2933 distinct citations, of which 2757 (94%) were excluded during the initial screening phase (Fig. 1). For the remaining 176 citations, full-text papers were ordered, obtained and



**Fig 1** Flow diagram of study selection.

a. Some studies excluded for multiple reasons.

independently assessed against the eligibility criteria, with five discrepancies resolved by discussion (97% agreement,  $k=0.75$ ). Nine studies met the inclusion criteria.<sup>16–24</sup> The main reasons for exclusion of full-text papers were use of non-randomised designs, primary end-points less than 3 months and depression not required for participant eligibility.

### Characteristics of included studies

Of the nine included trials (Table 1), four were conducted in the USA,<sup>16–19</sup> and one each in the UK,<sup>20</sup> Australia,<sup>21</sup> New Zealand,<sup>22</sup> China<sup>23</sup> and Hong Kong.<sup>24</sup> Three trials were explicitly identified in the study report as being either feasibility,<sup>21</sup> pilot<sup>16</sup> or efficacy studies.<sup>18</sup> The nine trials randomised 667 participants (69% female), with sample size ranging from 14 to 193. The mean age of trial populations ranged from 65 years<sup>20</sup> to over 80 years.<sup>19,22,24</sup>

Depression eligibility was determined by clinician diagnosis,<sup>17,18,20</sup> symptom checklist,<sup>16,19,21,23</sup> either a diagnosis or symptom checklist,<sup>24</sup> or a three-question depression screen validated for use in primary care.<sup>22</sup> Baseline receipt of anti-depressant medication was required for eligibility in one trial,<sup>20</sup> was an exclusion criterion in three<sup>16,18,21</sup> and allowed but not required in four.<sup>17,19,22,24</sup> Common exclusion criteria included medical conditions for which exercise was contraindicated, psychiatric illness, cognitive impairment, alcohol or substance misuse and, to a lesser extent, being a regular exerciser<sup>18,20</sup> or lacking motivation to exercise.<sup>21</sup>

In two trials the exercise intervention was classified as three-dimensional (3D) training, which included Tai Chi<sup>23</sup> and Qi Gong.<sup>24</sup> The remaining seven trials included elements of endurance and strength training, and were classified as mixed exercise, including four trials described in the trial's report as mixed, two<sup>18,21</sup> that were based mostly on strength training but included elements of endurance training, and one<sup>17</sup> in which endurance training was prescribed and strength training activities were encouraged. Interventions typically involved exercising for three to five, 30–45 min sessions per week for 3–4 months. Exercise was completed in participants' homes,<sup>17,22</sup> including care homes,<sup>19,24</sup> and various community-based facilities.<sup>16,18,20,21,23</sup>

**Table 1 Characteristics of included studies**

Study, author (country), depression eligibility	Participants, <i>n</i> (% female); age, years mean (s.d.); depression, mean (s.d.)	Interventions, exercise type, intensity, frequency and duration, programme length, setting and provider	Outcome, measure (end-point): mean (s.d.)
Brenes <sup>16</sup> (USA), PHQ-9 (2–4 symptoms)	37 (62); 74.6 (6.6); GDS-15: I = 7.0 (3.0), C = 7.8 (4.2)	I: Moderate-/high-intensity aerobic and resistance training at three 60 min weekly group sessions for 4 months. Delivered by ACSM instructor in local sports centre. C: Telephone discussion with researcher about health status at 2, 6, 10 and 14 weeks	GDS-15 (4 months): I = 4.5 (2.9), C = 6.3 (3.5)
Chiechanowski <sup>17</sup> (USA), DSM-IV minor depression or dysthymia	138 (79); 73.0 (8.5); HSCL-20: I = 1.3 (0.5), C = 1.2 (0.5)	I: Depression management programme of eight 50 min individual sessions over 19 weeks, promoting 30 min moderate intensity activity on 5 days per week. C: Physician informed of positive depression screen and advised usual care	HSCL-20 (12 months): I = 0.82 (0.62), C = 1.01 (0.46)
Chou <sup>23</sup> (China), CES-D $\geq$ 16	14 (50); 72.6 (4.2); CES-D: I = 32.0 (9.9), C = 32.7 (8.7)	I: Moderate-intensity Tai Chi at three 45 min weekly group sessions for 3 months. Delivered by Tai Chi practitioner in a psychogeriatric out-patient clinic. C: No intervention, waiting list	CES-D (3 months): I = 15.3 (9.8), C = 39.1 (9.7)
Kerse <sup>22</sup> (New Zealand), 3-question, primary care depression screen	193 (58.5); 81.1 (4.4); GDS-15: I = 3.4 (2.7), C = 4.0 (2.8)	I: Otago programme – moderate-intensity balance, progressive resistance and strengthening exercises and walking, each at three 30 min sessions per week for 6 months. Delivered at home by trained nurse in seven visits over 3 months. C: Equal social contact with nurse using conversational guide	GDS-15 (12 months): I = 2.4 (2.2), C = 2.8 (2.2)
Mather <sup>20</sup> (UK), ICD-10 mood disorder + GDS $\geq$ 10 + antidepressant	86 (69); 64.9 (range 53–91); HRSD: I = 16.7, C = 17.4	I: Mixed endurance, strength and stretching exercise at two 45 min group sessions per week for 10 weeks. C: Equal contact health education talks, including sessions on depression and exercise	HRSD (34 weeks): I = 11.5 (3.3), C = 13.7 (3.3)
Sims <sup>21</sup> (Australia), GDS $\geq$ 11	38 (66); 74.3 (5.9); GDS-30: I = 12.6 (3.6), C = 12.2 (3.5)	I: Moderate-intensity progressive resistance training at three 30 min weekly sessions for 10 weeks. Tailored and delivered by gym instructor in local gym. C: Exercise advice and information about local exercise options	GDS-30 (6 months): I = 11.5 (6.7), C = 11.9 (4.9)
Singh <sup>18</sup> (USA), DSM-IV criteria for (unipolar) depression including dysthymia	32 (63); 71 (2.0); BDI: I = 21.0 (2.0), C = 18.3 (1.8)	I: High-intensity progressive resistance training at three 45 min weekly group sessions for 10 weeks, supervised at university facility, followed by 10 weeks, unsupervised exercise at home with weekly telephone support. C: Health education lectures and videos at two 1 h group sessions per week for 10 weeks	BDI (20 weeks): I = 9.2 (2.8), C = 11.0 (2.4)
Tsang <sup>24</sup> (Hong Kong), current diagnosis or GDS (cut-off NR)	97 (81); 82.3 (NR); GDS-15: I = 5.17 (2.8), C = 6.50 (1.4)	I: Moderate-intensity Qi Gong at three 30–45 min weekly group sessions with 15 min/day unsupervised Qi Gong for 16 weeks. Delivered in care home by Qi Gong practitioner. C: Newspaper reading group with therapist in care home – equal contact	GDS-15 (24 weeks): I = 3.4 (2.5), C = 5.7 (1.5)
Williams <sup>19</sup> (USA), CSDD $\geq$ 7	32 (89); 87.9 (6.0); CSDD: I = 12.18 (5.0), C = 14.6 (5.8)	All interventions delivered by interventionists/researchers in care home at five 30 min individual weekly sessions for 16 weeks. I: Moderate-intensity mixed exercise – strength, balance, flexibility, walking. C: Social conversation	CSDD (16 weeks): I = 9.7 (6.6), C = 11.8 (8.1)

I, intervention group; C, control group; ACSM, American College of Sports Medicine; PHQ-9, Patient Health Questionnaire-9; GDS, Geriatric Depression Scale; HSCL-20, Hopkins Symptom Checklist-20; CES-D, Center for Epidemiologic Studies Depression Scale; HRSD, Hamilton Rating Scale for Depression; BDI, Beck Depression Inventory; CSDD: Cornell Scale for Depression in Dementia; NR, not reported.

Exercise was supervised in all but two trials<sup>17,22</sup> and completed in either group<sup>16,18,20,23,24</sup> or individual<sup>17,19,21,22</sup> formats.

Four trials compared exercise alone to an active usual care control, which included a referral letter sent to the primary care clinician recommending usual care,<sup>17</sup> brief advice about exercise,<sup>21</sup> a structured health education programme<sup>20</sup> and telephone discussions of health status.<sup>16</sup> A waiting-list control group (i.e. no contact intervention) was used in one trial,<sup>23</sup> whereas in four trials exercise was compared with a non-active control intervention, for example equal contact or attentional control.<sup>18,19,22,24</sup> Outcomes were assessed at follow-up ranging from 3 to 12 months. In four trials the primary end-point (3–6 months) coincided with the end of the intervention period<sup>16,18,19,23</sup> and in five it was 2–6 months post-intervention. In one study<sup>18</sup> outcomes were assessed at 20 weeks and at 26 months, but only data from the former are synthesised so as to avoid introducing between-study variation in follow-up assessments.

In four trials<sup>17,20–22</sup> that reported the number of eligible patients invited to participate in the trial, the uptake, or recruitment rate, was 52% (38 of 73),<sup>21</sup> 55% (193 of 353),<sup>22</sup> 76% (86 of 113)<sup>20</sup> and 92% (138 of 150).<sup>17</sup> Four trials<sup>18,20–22</sup> assessed adherence to the exercise intervention. In two trials, 75% (12/16)<sup>18</sup> and 58% (11/19)<sup>21</sup> of participants satisfied the adherence criterion of attendance at  $\geq$  20 of 30 exercise sessions.

One trial<sup>20</sup> reported that the mean attendance at exercise sessions was 67%, which approximates to 13 of 20 exercise sessions.<sup>20</sup> In the final trial,<sup>22</sup> adherence was defined as completing  $\geq$  2 of 3 prescribed exercise sessions per week as well as  $\geq$  2 of 3 recommended walking sessions per week. At 6 months (1 month post-intervention), 64% of participants satisfied the criteria for adherence and, at 12 months (7 months post-intervention), 57% were adherent.

Risk of bias varied among the included studies (Appendix 1). Risk was assessed as high in two trials,<sup>23,24</sup> moderate in four<sup>16,18,19,21</sup> and low in three.<sup>17,20,22</sup> Across the nine trials a total of 27 (60%) risk of bias items were assessed as adequate, 11 (24%) were unclear and 7 (16%) were inadequate. Common methodological limitations included failure to analyse data according to the intention-to-treat principle, lack of masked outcome assessment and incomplete follow-up of participants. Risk of bias assessment was hindered by poor reporting practices, including both inconsistent and insufficient reporting.

### Effect of exercise on depression

The point estimate of effect for each trial indicated lower depression severity among participants allocated to the exercise group compared with those allocated to the non-exercise control

(Fig. 2). In four<sup>17,20,23,24</sup> of nine trials the difference in depression severity was statistically significant. Two trials<sup>23,24</sup> of 3D exercise reported effect sizes of far greater magnitude than the remaining trials, and statistical heterogeneity was detected among trial-level effects ( $I^2 = 58\%$ ,  $\chi^2 = 18.97$ , d.f. = 8,  $P = 0.02$ ). The two trials were removed and neither contributed to subsequent analyses. The decision not to combine the trials in a separate synthesis was based on the detection of statistical heterogeneity between the trials ( $I^2 = 60\%$ ,  $\chi^2 = 2.48$ , d.f. = 1,  $P = 0.12$ ) and assessment of each trial as having high risk of bias.

Only trials of mixed exercise contributed data to the pooled analyses (Table 2). The synthesis of data from seven trials produced a small but statistically significant effect in which exercise was associated with lower severity of depression (SMD =  $-0.34$ , 95% CI  $-0.52$  to  $-0.17$ ). There was no evidence of statistical heterogeneity among the pooled estimates ( $I^2 = 0\%$ ), and no indication of small study bias (Egger  $-0.52$ , 95% CI  $-3.72$  to  $2.69$ ,  $P = 0.72$ ).

Small, statistically significant effects emerged from the synthesis of three trials<sup>17,18,20</sup> in which participant eligibility required a current diagnosis of depression (SMD =  $-0.38$ , 95% CI  $-0.67$  to  $-0.10$ ), and in four trials<sup>16,19,21,22</sup> using a symptom checklist threshold (SMD =  $-0.34$ , 95% CI  $-0.62$  to  $-0.06$ ). For the latter synthesis there was some indication of variation among the pooled estimates ( $I^2 = 25\%$ ), but this was unlikely to be important and did not exceed what would be expected by chance alone ( $\chi^2 = 4.02$ , d.f. = 3,  $P = 0.26$ ).

Small, statistically significant effects favouring exercise were observed in three trials<sup>17,20,22</sup> with low risk of bias (SMD =  $-0.36$ , 95% CI  $-0.61$  to  $-0.10$ ), four trials<sup>17,18,20,21</sup> using an active intervention control (SMD =  $-0.44$ , 95% CI  $-0.67$  to  $-0.20$ ), and four trials<sup>17,20-22</sup> with extended follow-up (SMD =  $-0.32$ , 95% CI  $-0.54$  to  $-0.10$ ). Variation among

pooled estimates was detected but did not exceed what would be expected by chance alone in the analyses for risk of bias ( $I^2 = 25\%$ ,  $\chi^2 = 3.20$ , d.f. = 2;  $P = 0.20$ ) and follow-up period ( $I^2 = 19\%$ ,  $\chi^2 = 3.72$ ; d.f. = 3,  $P = 0.29$ ). There was no evidence of statistical heterogeneity among trials comparing exercise with an active usual care control ( $I^2 = 0\%$ ,  $\chi^2 = 2.43$ ; d.f. = 3,  $P = 0.49$ ).

## Discussion

### Summary of main results

The review identified nine RCTs evaluating the medium-term (3–12 months) effect of exercise on the severity of depressive symptoms in older people. Synthesised data from seven trials of mixed exercise indicated a small but statistically significant effect favouring exercise. Small, statistically significant effects favouring exercise were similarly observed in a pre-planned analysis of trials stratified by depression eligibility criteria (clinician diagnosis or symptom checklist threshold). These findings were robust in sensitivity analyses that excluded trials with higher risk of bias, non-active intervention comparison groups or in which the primary end-point was within rather than beyond the intervention period.

### Strengths and weaknesses of the study

The study adhered to the pre-specified protocol, adopted procedures to limit the potential for bias and used appropriate methods to select, evaluate and synthesise relevant evidence. A comprehensive search for published and unpublished studies, which included multiple electronic databases and scanning of bibliographies, yielded nine trials, all of which were published studies. Absence of data from unpublished studies is a potential weakness, since effects estimated from published studies may be

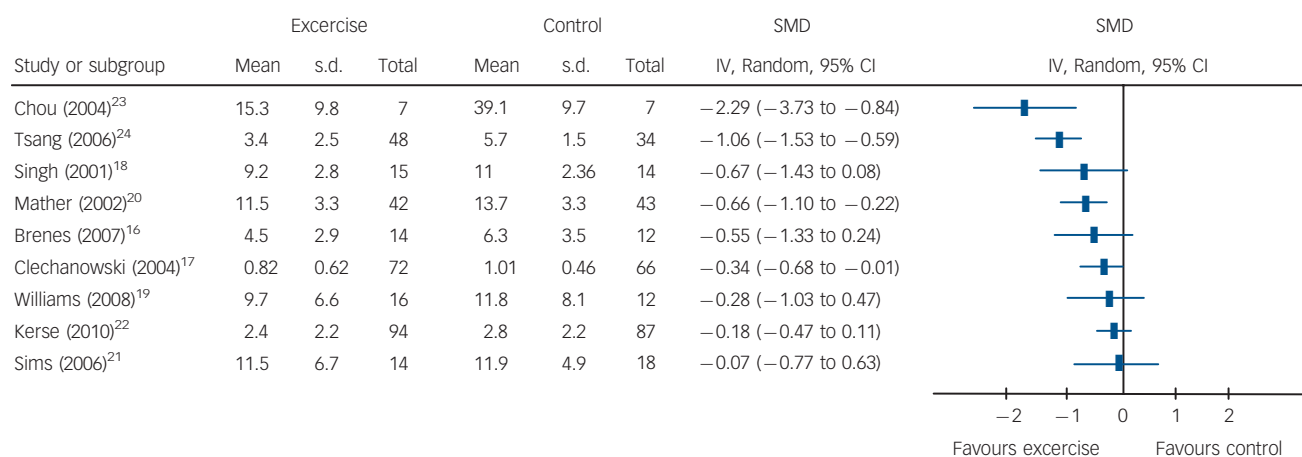


Fig. 2 Trial-level data, effect estimates and forest plots for depression severity. SMD, standard mean difference.

Table 2 Summary results for pooled analyses

Effect foci	Trials, <i>n</i> (participants, <i>n</i> )	Effect size, SMD (95% CI)	$I^2$ , %
Mixed exercise	7 (519)	$-0.34$ ( $-0.52$ to $-0.17$ )	0
Diagnostic criteria	3 (195)	$-0.38$ ( $-0.67$ to $-0.10$ )	0
Symptom checklist	4 (324)	$-0.34$ ( $-0.62$ to $-0.06$ )	25
Lowest risk of bias	3 (404)	$-0.36$ ( $-0.61$ to $-0.10$ )	27
Active control	4 (284)	$-0.44$ ( $-0.67$ to $-0.20$ )	0
Extended follow-up	4 (436)	$-0.32$ ( $-0.54$ to $-0.10$ )	19

SMD, standard mean difference.

inflated because of bias towards the non-publication of small studies with null effects. However, null effects on the outcome of interest in five of nine included trials, and five of seven synthesised trials, mitigates concerns about publication bias, since decisions to publish appear independent of the observed effect.

There was at least a moderate risk of bias in all but three of the included trials. As study quality and effect size typically show an inverse association, the underlying risk of bias may have inflated our estimate of the treatment effect. Sensitivity analysis restricted to three trials of low risk of bias yielded a pooled estimate of a nearly identical magnitude and precision as the estimate derived from synthesis that included trials of higher risk of bias. These data are inconsistent with the suggestion that bias, due to poor methodological quality, may have inflated the observed effect of mixed exercise on symptoms of depression.

A strength of the review is that it not only provides data crucial to healthcare decision-making, such as uptake of and adherence to exercise among the target population, but that data are derived from trials conducted under conditions that most closely match the context of usual healthcare practice. Specifically, in three trials of low risk of bias, with older people who were not excluded for or prohibited from use of antidepressant medication, who were identified, approached and invited to participate in the context of routine clinical practice, 68% (417 of 616) of eligible patients agreed to participate in a trial of exercise for treatment of depression, at least three-quarters of whom achieved the minimum criteria for adherence.

### Comparison with other studies

Other reviews of exercise for depression in older people<sup>11,12</sup> have included both randomised and non-randomised study designs, and trials in which current depression was not required for participant eligibility. Findings based on different levels of evidence in clinically heterogeneous populations are not easy to use to inform healthcare decisions. Moreover, previous reviews have used qualitative methods of synthesis and relied on quasi-quantitative methods for interpretation based on a simple count of studies with/without significant results. This approach is less than ideal, not least because there is an increased potential to conclude that exercise is beneficial for depression, when the magnitude of the effect is too small to be meaningful. Our study provides the first quantitative estimate of the effect of exercise on depression severity among older people with clinically significant symptoms of depression, and pre-planned subgroup and sensitivity analyses suggest that the effect is both stable and robust.

The pooled effect of exercise on depression severity observed in this review (SMD = -0.34) is comparable with the range of effects estimated for different classes of antidepressant medication (SMD = 0.2–0.5)<sup>25</sup> and psychotherapy (SMD = 0.18–0.34).<sup>26</sup> However, although age-associated factors can complicate use of antidepressant medication and resource-related factors can impede timely access to psychotherapy, for older people with or without medical morbidity, individualised mixed exercise has very few risks, is easy to access and has the potential to improve a wide range of additional health outcomes.

### Meaning and implications of the study

The clinical relevance of an SMD can more easily be considered when converted back into units of the original scale, or when represented as the overlap of distributions. At a group level, an SMD of -0.34 is equivalent to 63% of exercise participants having lower severity of depression than the average control participant or, put another way, 13% of the population of exercise

participants doing better than would otherwise have been expected. For individuals at the symptom checklist threshold, an SMD of -0.34 translates into a reduction of approximately 20% in the severity of depressive symptoms. The magnitude of effect estimated in this study is clinically meaningful at the individual level, and may have substantial public health significance at the population level.

Our findings must be interpreted in relation to the quantity and quality of available evidence. For exercise interventions involving 3D training (Tai Chi and Qi Gong), two trials with a high risk of bias demonstrate clearly that evidence was insufficient in both quantity and quality. For interventions involving mixed exercise, the available evidence comprised seven trials with low to moderate risk of bias. Although the quantity and quality of evidence was less than ideal, these limitations are not sufficient to dismiss the findings of the review. Evidence is drawn from RCTs of direct relevance to the population, intervention and outcome of interest. All analyses were pre-specified, synthesised results, yielded consistent effects and there was no evidence of small study effects, including publication bias. Thus, there is a moderate-quality evidence base for the medium-term effect of mixed exercise on depression severity.

The finding that mixed exercise has a small, but clinically important effect on symptoms of depression, has general applicability to people aged over 60 years who are experiencing elevated symptoms of depression. However, as depression may reduce the appeal of exercise, participants in exercise trials may not be representative of the population of older people with depression. As none of the included trials stratified randomisation by depression severity, it is unclear whether our findings are equally applicable to patients with elevated, but subthreshold, symptoms as they are to patients with more severe symptoms, such as those that satisfy diagnostic criteria. Similarly, the findings may have limited applicability for patients who are more frequent exercisers or who have more severe comorbid physical illness, since several trials excluded patients classified as regular exercisers or as too ill to participate.

Research to reduce residual uncertainty concerning the applicability of moderate-quality evidence should be considered a public health priority. This research should be in the form of a pragmatic RCT with sufficient power to detect an effect equivalent to an SMD of at least 0.3. Such research might usefully stratify randomisation by depression severity, receipt of antidepressant medication and/or level of regular exercise. As uptake of exercise in this population will be the crucial driver for cost-effectiveness, interventions should include integrated strategies, based on behaviour change techniques, to maximise uptake of and adherence to exercise regimens.

The findings of this review are consistent with the suggestion that, for older people who present with clinically meaningful symptoms of depression, prescribing structured exercise with mixed elements of endurance and strength training tailored to individual ability, will likely reduce the severity of depression. Whereas the evidence on the effect of mixed exercise is minimally sufficient, for Tai Chi and Qi Gong the available evidence is insufficient in both quantity and quality.

**Christopher Bridle**, PhD, **Kathleen Spanjers**, BSc, **Shilpa Patel**, CPsychol, **Nicola M. Atherton**, BSc, **Sarah E. Lamb**, DPhil, Warwick Clinical Trials Unit, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry, UK

**Correspondence:** Christopher Bridle, Warwick Clinical Trials Unit, Division of Health Sciences, Warwick Medical School, The University of Warwick, Coventry CV4 7AL, UK. Email: c.bridle@warwick.ac.uk

First received 4 Apr 2011, final revision 20 Nov 2011, accepted 17 Jan 2012

## Appendix

## Risk of bias within trials

Trial	Risk of bias item <sup>a</sup>					Overall risk <sup>b</sup>
	Sequence	Allocation	Masking	Follow-up	ITT	
Brenes (2007) <sup>16</sup>	✓	✓	✓	?	?	Moderate
Chiechanowski (2004) <sup>17</sup>	✓	✓	✓	✓	✓	Low
Chou (2004) <sup>23</sup>	?	?	×	?	×	High
Kerse (2010) <sup>22</sup>	✓	✓	✓	✓	✓	Low
Mather (2002) <sup>20</sup>	✓	✓	✓	✓	✓	Low
Sims (2006) <sup>21</sup>	✓	✓	×	✓	✓	Moderate
Singh (2001) <sup>18</sup>	✓	✓	×	✓	?	Moderate
Tsang (2006) <sup>24</sup>	?	?	?	×	×	High
Williams (2008) <sup>19</sup>	✓	?	✓	×	✓	Moderate

a. ✓, adequate; ?, unclear; ×, inadequate.  
b. Low risk: 5 items adequate; moderate: ≥3 items adequate and <2 items inadequate; high: ≥2 inadequate.

## References

- Alexopoulos GS. Depression in the elderly. *Lancet* 2005; **365**: 1961–70.
- Blazer DG. Depression in late life: review and commentary. *J Gerontol A Biol Sci Med Sci* 2003; **58**: 249–65.
- Katon WJ, Lin E, Russo J, Unützer J. Increased medical costs of a population-based sample of depressed elderly patients. *Arch Gen Psychiatry* 2003; **60**: 897–903.
- Birrer RM, Vemuri S. Depression in late life: a diagnostic and therapeutic challenge. *Am Fam Physician* 2004; **69**: 2375–82.
- Fernández A, Haro JM, Martínez-Alonso M, Demyttenaere K, Brugha TS, Autonell J, et al. Treatment adequacy for anxiety and depressive disorders in six European countries. *Br J Psychiatry* 2007; **190**: 172–3.
- Zivin K, Kales HC. Adherence to depression treatment in older adults: a narrative review. *Drugs Aging* 2008; **25**: 559–71.
- Goodwin RD. Association between physical activity and mental disorder among adults in the United States. *Prev Med* 2003; **36**: 698–703.
- Office for National Statistics. *Subnational Population Projections for England*. ONS, 2010 (<http://www.statistics.gov.uk/>).
- Barbour KA, Blumenthal JA. Exercise training and depression in older adults. *Neurobiol Aging* 2005; **26** (suppl 1): 119–23.
- Trivedi MH, Greer TL, Grannemann BD, Chambliss HO, Jordan AN. Exercise as an augmentation strategy for treatment of major depression. *J Psychiatr Pract* 2006; **12**: 205–13.
- Sjösten N, Kivelä SL. The effects of physical exercise on depressive symptoms among the aged: a systematic review. *Int J Geriatr Psychiatry* 2006; **21**: 410–8.
- Blake H, Mo P, Malik S, Thomas S. How effective are physical activity interventions for alleviating depressive symptoms in older people? A systematic review. *Clin Rehabil* 2009; **23**: 873–87.
- World Health Organization. *Ageing*. WHO, no date (<http://www.who.int/topics/ageing/en/>).
- Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions (V5.0.2)*. The Cochrane Collaboration, 2009.
- Hedges LV, Vevea JL. Fixed and random effects models in meta-analysis. *Psychol Methods* 1998; **3**: 486–504.
- Brenes GA, Williamson JD, Messier SP, Rejeski WJ, Pahor M, Ip E, et al. Treatment of minor depression in older adults: a pilot study comparing sertraline and exercise. *Aging Ment Health* 2007; **11**: 61–8.
- Ciechanowski P, Wagner E, Schmalting K, Schwartz S, Williams B, Diehr P, et al. Community-integrated home-based depression treatment in older adults: a randomized controlled trial. *JAMA* 2004; **291**: 1569–77.
- Singh NA, Clements KM, Singh MAF. The efficacy of exercise as a long-term antidepressant in elderly subjects: a randomized, controlled trial. *J Gerontol A Biol* 2001; **56**: 479–501.
- Williams CL, Tappen RM. Exercise training for depressed older adults with Alzheimer's disease. *Aging Ment Health* 2008; **12**: 72–80.
- Mather AS, Rodriguez C, Guthrie MF, McHarg AM, Reid IC, McMurdo MET. Effects of exercise on depressive symptoms in older adults with poorly responsive depressive disorder. Randomised controlled trial. *Br J Psychiatry* 2002; **180**: 411–5.
- Sims J, Hill K, Davidson S, Gunn J, Huang N. Exploring the feasibility of a community-based strength training program for older people with depressive symptoms and its impact on depressive symptoms. *BMC Geriatr* 2006; **6**: 18.
- Kerse N, Hayman KJ, Moyes SA, Peri K, Robinson E, Dowell A, et al. Home-based activity programme for older people with depressive symptoms: DeLLITE – a randomised controlled trial. *Ann Fam Med* 2010; **8**: 214–23.
- Chou KL, Lee PWH, Yu ECS, Yeung-Hung Chen DM, Chan SSC, Chi I. Effect of Tai Chi on depressive symptoms amongst Chinese older patients with depressive disorders: a randomized clinical trial. *Int J Geriatr Psychiatry* 2004; **19**: 1105–7.
- Tsang HW, Fung KM, Chan AS, Lee G, Chan F. Effect of a Qi Gong exercise programme on elderly with depression. *Int J Geriatr Psychiatry* 2006; **21**: 890–7.
- Taylor D, Meader N, Bird V, Pilling S, Creed F, Goldberg D. Pharmacological interventions for people with depression and chronic physical health problems: systematic review and meta-analyses of safety and efficacy. *Br J Psychiatry* 2011; **198**: 179–88.
- Cuijpers P, van Straten A, Bohlmeijer E, Hollon D, Anderson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med* 2010; **40**: 211–23.

