Comparative Effectiveness of Three Exercise Types to Treat Clinical Depression in Older Adults: A Systematic Review and Network Meta-Analysis of Randomised Controlled Trials

Kyle J. Miller\textsuperscript{a}, Daniela C. Gonçalves-Bradley\textsuperscript{b}, Pinyadapat Areerob\textsuperscript{a}, Declan Hennessy\textsuperscript{a}, Christopher Mesagno\textsuperscript{a}, & Fergal Grace\textsuperscript{a}

\textsuperscript{a}School of Health and Life Sciences, Federation University, Ballarat, Victoria, Australia
\textsuperscript{b}Nuffield Department of Population Health, University of Oxford, UK

\textbf{Corresponding author:} Kyle Miller (kj.miller@federation.edu.au)

\textbf{Declarations of Interest:} None

\textbf{Funding:} Kyle Miller was supported by an Australian Government Research Training Program (RTP) Fee-Offset Scholarship through Federation University Australia.
Abstract

Background

Few studies have directly compared the effects of different exercise therapies on clinical depression in older adults. Thus, we conducted a systematic review and network meta-analysis of current evidence from randomised controlled trials (RCTs) to compare the effectiveness of three major exercise types (aerobic, resistance, and mind-body exercise) in clinically depressed older adults.

Methods

We followed PRISMA-NMA guidelines and searched databases for eligible RCTs (inception – September 12th, 2019). RCTs were eligible if they included clinically depressed adults aged >65 years, implemented one or more exercise therapy arms using aerobic, resistance, or mind-body exercise, and assessed depressive symptoms at baseline and follow-up using a validated clinical questionnaire.

Results

A network meta-analysis was performed on 15 eligible RCTs comprising 596 participants (321 treatment and 275 controls), including aerobic (n = 6), resistance (n = 5), and mind-body (n = 4) exercise trials. Compared with controls, mind-body exercise showed the largest improvement on depressive symptoms (g = -0.87 to -1.38), followed by aerobic exercise (g = -0.51 to -1.02), and resistance exercise (g = -0.41 to -0.92). Notably, there were no statistically significant differences between exercise types: aerobic versus resistance (g = -0.10, PrI = -2.23, 2.03), mind-body versus aerobic (g = -0.36, PrI = -2.69, 1.97), or mind-body versus resistance (g = -0.46, PrI = -2.75, 1.83).

Conclusions
These findings should guide optimal exercise prescription for allied health professionals and stakeholders in clinical geriatrics. Notably, clinically depressed older adults may be encouraged to self-select their preferred exercise type in order to achieve therapeutic benefit on symptoms of depression. In coalition with high levels of compliance, these data provide encouraging evidence for the antidepressant effect of either aerobic, resistance, or mind-body exercise as effective treatment adjuncts for older adults presenting with clinical depression.

**Keywords**

RCT; physical activity; aerobic; resistance; mind-body; major depressive disorder
1. Introduction

Population studies consistently highlight that older adults aged 65 years and over have an elevated risk of presenting with acute and chronic depression, estimated to exceed 15-20% in some cohorts (Luppa et al., 2012; Seitz et al. 2010; Volkert et al., 2013). Clinical depression during older age is linked to higher rates of physical illness and disease (Sinnige et al., 2013), disability and cognitive impairment (Scurti et al., 2011), and comorbid psychiatric disorders (Laborde et al., 2014). Therefore, it is unsurprising that an estimated 20-50% of depressed older adults have a poor prognosis or are prone to relapse (Comijs et al., 2015; Licht-Strunk et al., 2007). In light of the projections in ageing demographics and impending demand on healthcare systems, there is shared motivation to optimise clinically meaningful, non-pharmacological adjunctive therapies that are allied to the medical treatment of age-related challenges to mental well-being.

Older adults are more reluctant to discuss their mental health and seek necessary treatment for clinical depression than their younger counterparts (Garrido et al., 2011; Wang et al., 2005). To compound this issue, there are unique obstacles to successful treatment, such as the widely acknowledged phenomenon whereby a significant proportion (~30%) of older adults have problematic compliance with their antidepressant medication (Rossom et al., 2016; Stein-Shvachman et al., 2013), as well as non-compliance often leading to worsening of symptoms and other age-related comorbidities (Ho et al., 2016). Pharmacological and psychotherapeutic antidepressant treatments, either alone or in combination, are the \textit{sine qua non} for standard medical treatment of clinical depression in older adults. However, this does not forego the requirement for a broad appreciation of the adjunctive non-pharmacological treatments for
clinical depression in optimising mental health-span and supporting the treatment of clinical depression during older age.

There is broad cross-disciplinary acknowledgement amongst medical and allied health professions that physical exercise has the potential for clinically meaningful antidepressant effects. We can generalise exercise into three major exercise types: aerobic, resistance, and mind-body exercise (American Psychiatric Association, 2010; National Health, Lung, and Blood Institute 2006; U.S. Department of Health and Human Services, 2018). Patrons of accumulated research knowledge declare that routine engagement in physical activity and/or exercise has the potential to impact symptoms of clinical depression (American College of Sports Medicine, 2017; American Psychiatric Association, 2010; World Health Organization, 2010). It may, therefore, seem surprising that there has been little attempt to compare the effectiveness of different exercise modes to gain a deeper understanding of exercise as a bona fide adjunctive therapy. Indeed, ongoing public health branding strategies, such as ‘Exercise is Medicine’, are promulgated to connect a broad message towards the pleiotropic benefits of physical activity to improve characteristics such as mobility, cardio-respiratory fitness, muscular strength, and body-mindfulness. However, there are profound physiological, mechanical, and metabolic differences between the three broad exercise modes (aerobic, resistance, and mind-body) and few data on the comparative effectiveness to treat symptoms of clinical depression in older adults.

Aerobic exercise is a distinct form of activity that involves the integration of large muscle groups, such as in the rhythmic propulsion of body mass during movements of different intensities (e.g., walking, jogging, or running) or activities of lower mechanical impact (e.g., swimming or cycling). Aerobic exercise is matched by increasing metabolic effort, which necessitates effort of breathing, heart rate, and blood flow to match the effort of intensity
American College of Sports Medicine, 2017; Nelson et al., 2007). In contrast, resistance exercise refers to a specialised method of muscular conditioning which normally requires a variety of equipment, such as free weights, weight machines, elastic bands, and/or body weight exercises (Faigenbaum et al., 2009). Resistance exercise is distinct from aerobic or mind-body exercise in that there is a focus on repeatedly overloading muscles during static, isometric, or dynamic contractions, which cause a closer and stronger connection between the musculature and nervous system. Remarkably, only one randomised controlled trial (RCT) has directly compared the effectiveness of aerobic versus resistance exercise in clinically depressed older adults (Penninx et al., 2002). A subsequent analysis of this 18-month Fitness, Arthritis and Seniors Trial (Ettinger et al., 1997) revealed that aerobic exercise induced a more favourable antidepressant effect than resistance exercise, highlighting that there may be distinguishable differences between aerobic and resistance exercise types (Penninx et al., 2002).

Mind-body exercise incorporates a range of low impact and deliberately slow movements in addition to breathing, meditation, and/or progressive relaxation. This exercise mode integrates low-intensity muscular activity, such as flexibility or balance training, with an internally directed focus that encourages a self-contemplative mental state (La Forge, 1997). Many forms of mind-body exercise are available for older populations (e.g., yoga, tai chi, qigong), yet a single RCT has directly compared the effectiveness of mind-body exercise with other exercise treatments. Prakhinit et al. (2014) compared the effects of a traditional aerobic walking program with a Buddhism-based walking meditation program in a sample of 45 depressed elderly. The walking meditation program resulted in a significantly larger decrease in depression than aerobic walking alone, suggesting that exercise incorporating both mind and body components are most effective for elderly with mild and moderate depressive symptoms (Prakhinit et al., 2014).
Perhaps the foremost obstacle to exercise prescription for mental health during ageing is that there are few adequately controlled trials that compare different exercise types within the same investigation. To approach this important research problem, recent meta-analytical reviews have been published to indirectly compare the antidepressive effects of different exercise types by employing subgroup analysis of pooled literature (Bridle et al., 2012; Heinzel et al., 2015; Schuch et al., 2016), resulting in some interesting findings. For instance, interventions using a combination of aerobic and resistance exercises appear to be most effective for the treatment of depressive symptoms in clinically depressed older adults (Bridle et al., 2012; Schuch et al., 2016). There is, however, some discordance in these efforts to compare exercise modalities, as Heinzel and colleagues (2015) concluded that mind-body exercise is most effective out of the three types of exercise.

There is coalition between the known reluctance of older adults to participate in formal exercise regimens, as well as confusion of specific knowledge amongst clinicians, that restricts clinical guidance on exercise prescription. Clarity on the comparative effectiveness of the antidepressive effects of different exercise modalities might inform clinician-patient discussion, as well as encourage patients to have more control over their selection of the best evidence-based therapeutic exercise that is allied to ongoing treatment for clinical depression. Meta-analysis, however, has an inherent lack of precision to accurately examine the comparative effectiveness of related exercise therapies, which is required to better understand the antidepressive effects of separate exercise types. Network meta-analysis offers a solution to this problem by statistically comparing treatment effects by simultaneously assessing both the direct and indirect evidence of independent treatments. This goes beyond simple pairwise comparisons and provides a more
accurate calculation of true effect size estimates than would otherwise be achievable with meta-
analysis using subgroups (Caldwell et al., 2005; Salanti et al., 2008).

Thus, the primary purpose of this review was to perform a systematic review and network
meta-analysis of current evidence from RCTs to compare the effectiveness of three major
exercise therapies (aerobic, resistance, or mind-body exercise) in clinically depressed older
adults. We hypothesised that (1) aerobic, resistance, and mind-body exercise would have larger
pooled effects on depressive symptoms than wait-list, usual care, and attention-control
comparisons, and (2) mind-body exercise would have a larger pooled effect on depressive
symptoms than either aerobic or resistance exercise. Secondary outcomes including adverse
events, adherence, and attrition were used to evaluate compliance.

2. Methods

We prospectively registered the current review in the PROSPERO database (registration
number: CRD42018094667). Both the network meta-analysis extension for the Preferred
Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA-NMA) guidelines
(Hutton et al., 2015) and the Cochrane Intervention Review that Compares Multiple
Interventions (Cochrane Methods, 2014) were used as a guide for this review. Specific geriatric
guidelines for meta-analyses (Shenkin et al., 2017) were also consulted to identify
inclusion/exclusion criteria, effect modifiers, and potential risks of bias.

2.1. Eligibility Criteria

We restricted studies to those where (1) participants were diagnosed with clinical
depression at baseline according to a structured diagnostic interview based on Diagnostic and
Statistical Manual of Mental Disorders (DSM) or International Statistical Classification of
Diseases and Related Health Problems (ICD) criteria, or a clinical threshold on a questionnaire validated against a structured diagnostic interview, (2) an RCT protocol was used to assess the effectiveness of one or more exercise therapy arms, (3) depressive symptoms were assessed at baseline and follow-up using a questionnaire that had been psychometrically validated against DSM or ICD criteria, and (4) a minimum mean sample age of 65 years was included. We excluded studies if (1) participants were not clinically depressed at baseline, (2) the intervention condition used a multicomponent treatment with non-exercise components, (3) the intervention condition could not be exclusively categorised as one of three types of exercise (aerobic, resistance, or mind-body), (4) depressive symptoms were not assessed as an outcome, based on change scores between baseline and follow-up, or (5) the sample had a mean age below 65 years.

Clinical cut-off scores on questionnaires validated against a structured diagnostic interview based on DSM or ICD criteria were used to determine if participants were diagnosed with depression at baseline. These included scores of 10 or higher on the Beck Depression Inventory (BDI; Beck et al., 1988, 1996), 16 or higher on the Center for Epidemiological Studies Depression Scale (CESD; Lewinsohn et al., 1997; Radloff, 1977), six or higher on the Cornell Scale for Depression in Dementia (CSDD; Kørner et al., 2006), nine or higher on the Geriatric Depression Scale-30 (GDS-30; Kørner et al., 2006; Wancata et al., 2006), five or higher on the Geriatric Depression Scale-15 (GDS-15; Almeida & Almeida, 1999; Kørner et al., 2006; Wancata et al., 2006), and 12 or higher on the Hamilton Rating Scale for Depression (HRSD; Baghy et al., 2004).

2.2. Literature Search

PubMed, Web of Science, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Health Source: Nursing/Academic Edition, PsycARTICLES, PsycINFO, and
SPORTDiscus were searched from the date of their inception up to September 12th, 2019. Text mining procedures were used to identify key search terms related to population, intervention, and study design. Full search terms and an example of the search strategy can be found in the Supplementary File S1. References lists from included studies, systematic reviews, and meta-analyses (i.e., Blake et al., 2009; Bridle et al., 2012; Heinzel et al., 2015; Mura & Carta, 2013; Rhyner & Watts, 2016; Schuch et al., 2016; Sjösten & Kivelä, 2006) were screened for additional articles.

2.3. Data Extraction and Coding

Study characteristics and outcome statistics were extracted independently by two researchers (KJM, PA, and/or DH) using a data extraction form (see Supplementary File S1) and discrepancies were resolved by consensus with a fourth researcher (CM). Authors of particular studies were contacted if a sufficiently detailed report of the data was not available. Type of control group was categorised as wait-list, usual control, or attention-control. Wait-list conditions were assigned to a waiting list and received the exercise intervention once the trial had concluded. Usual care conditions had no mention of a waiting list or any additional activities during the trial. Attention-control conditions (also known as attention placebo control or active control) included the engagement in a non-physical activity during the trial (e.g., social activities, educational programs, etc.).

Study characteristics were used to evaluate the plausibility of the transitivity assumption and evaluate the quality of included studies (see Table 1). Methodological characteristics were coded according to the intention-to-treat principle, use of a cluster design, and the depression measure(s) used. Participant characteristics were coded according to sample size, age (mean and standard deviation), percentage of females, depression diagnosis, and primary place of residence.
Type of exercise was categorised as aerobic, resistance, or mind-body exercise therapy. Length of the program was coded as total number of weeks. Intensity was evaluated based on ratings of perceived exertion (Borg, 1982), heart rate maximum (HR\text{max}; low = <50%, moderate = 50-70%, high = >70%), maximal oxygen uptake (VO\text{2max}; low = <40%, moderate = 40-60%, high = >60%) or one-repetition maximum (1R\text{max}; low = <50%, moderate = 50-74%, high = >74%; Sigal et al., 2004). Frequency was coded as total sessions per week. Duration was coded as the average number of minutes per session. Format of program was dichotomously coded as group or individual training. Format of supervision was dichotomously coded as supervised or unsupervised training. Agreement between the three researchers was 91.3%.

2.4. Risk of Bias and Quality Assessment

We assessed the risk of bias in the included studies using the Cochrane Collaboration’s Tool for Assessing Risk of Bias (Higgins et al., 2011). Judgements for ‘other sources of bias’ were based on small sample size (n < 15), low adherence (less than 80%), cluster randomisation, and inequity in the selection of the participants. Appraisal was performed by two researchers (KJM, PA, and/or DH), and consensus was reached with a fourth researcher (CM).

2.5. Summary of Outcomes

Primary outcome statistics including means (M), standard deviations (SD), and sample sizes (n) were used to calculate the mean changes in depressive symptoms. If a study reported more than one post-treatment depression score (e.g., midway, follow-up), only the assessment time-point immediately following the conclusion of the intervention phase was used. If a study reported depression scores on multiple outcome measures, only the most clinically relevant
depression measure was used (Costa et al., 2016; Smarr & Keefer, 2011). Pairwise relative
treatment effects for depressive symptoms were estimated using Hedges’ $g$ (Hedges, 1981) and
the 95% confidence interval (95% CI). Hedges’ $g$ coefficients were interpreted according to
Cohen (1988) conversions, whereby effects were considered small (0.2), medium (0.5), and large
(0.8). Secondary outcome data were also extracted for adverse events, adherence, and attrition.
Adverse events were qualitatively reported according to the descriptive information available in
the transcript. Attrition and adherence were reported as the percentage of total
attendance/attrition during the exercise intervention.

2.6. Data Synthesis

An underlying assumption of network meta-analysis is that estimates are derived from a
pool of studies with homogenous between-study effect modifiers (Jansen & Naci, 2013). If the
distribution of an effect modifier is heterogeneous across studies, the assumption of transitivity
may be violated. For instance, if no distinction is made between different control conditions,
exercise conditions weighted heavily on a single control comparison may artificially inflate or
deflate effect size estimates (Pagoto et al., 2013). Within the context of the current network
meta-analysis, exercise conditions (i.e., aerobic, resistance, mind-body) and control conditions
(i.e., wait-list, usual care, attention-control) were separated into separate network estimates to
account for the between-study variation of these effect modifiers.

Data were analysed and figures were generated using STATA/SE 15.1 (StataCorp, 2017).
Further description of the statistical procedures can be found in the Supplementary File S1. We
evaluated publication bias and small-study effects using a comparison-adjusted funnel plot. The
assumption of transitivity was assessed with exploratory meta-regression sensitivity analyses
comparing the distribution of participant and exercise modifiers across treatment comparisons.
Participant, intervention, and methodological characteristics were assessed, which have been identified as potential risks to transitivity in geriatric meta-analyses (Netz et al., 2005; Rhyner & Watts, 2016; Schuch et al., 2016).

A multivariate random-effects meta-analysis was performed using the ‘mvmeta’ command (White, 2009). A random-effects model assumes variance both within and between studies, explaining the heterogeneity of treatment effects (Borenstein et al., 2010). A common heterogeneity parameter was assumed across comparisons. Heterogeneity was evaluated using tau-squared ($\tau^2$), which estimates the deviation in effect sizes across the population of studies (Borenstein et al., 2009). The 95% prediction interval ($PrI$) was used to estimate the true dispersion of effect within two standard deviations of the mean effect size (Borenstein et al., 2009). The certainty of evidence contributing to network estimates was assessed with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach (Salanti et al., 2014). The significance level was $p < .05$ for all analyses.

3. Results

3.1. Study Selection

The literature search yielded a total of 2,309 eligible records after duplicate studies were removed ($n = 1,395$). Titles and abstracts were screened by two independent researchers (KJM and DCG) with an agreement of 91.9%. We assessed 356 full-text articles for eligibility, of which 340 were excluded from the systematic review (see Figure 1). Notably, one study (Shahidi et al., 2011) fulfilled all criteria but was excluded from the quantitative analysis because the control condition was not clearly defined, and the authors were unobtainable. The remaining 15 studies formed the data used in the final analyses. Full-text screening was performed
independently by two researchers (KJM and PA) with an agreement of 81.8% and discrepancies were resolved by consensus from a third researcher (CM).

[Figure 1 Approximately Here]

Figure 1. Flowchart of screening process.

3.2. Characteristics of the Studies

Data from a total of 596 participants (321 treatment and 275 controls) across 15 studies (aerobic = 6, resistance = 5, mind-body = 4) were included in the network meta-analysis (see Figure 2 for a network plot of pairwise comparisons across all studies). In general, 10 samples included participants who were young-old (65-75 years) and five included participants who were old-old (>75 years), with one sample uncategorised (>65 years). Participants were identified as either community-dwelling (n = 9) or living in residential care (n = 6). Samples were typically healthy, except for one sample including participants with a history of stroke (Sims et al., 2009) and another including participants with chronic medical illnesses (Tsang et al., 2013). An additional two RCTs included participants with dementia (Cheng et al., 2012; Williams & Tappen, 2008). Only five trials included participants with sedentary or low activity lifestyles (Chou et al., 2004; Favilla, 1992; Singh et al., 2005). Antidepressant medication was not adequately reported and/or controlled in most studies. Five studies excluded participants if they were currently using antidepressants (Sims et al., 2006; Singh et al., 1997, 2005), or unless the medication was stabilised for at least 2-3 months (Favilla, 1992; Haboush et al., 2006). Only three studies reported the use of antidepressants, comprising of 11% (Favilla, 1992), 17% (Cheng et al., 2012), and 36% (Williams & Tappen, 2008) of the total samples.
Five out of six aerobic RCTs included walking interventions (Lok et al., 2017; McNeil et al., 1991; Patel, 2004; Prakhinkit et al., 2014; Williams & Tappen, 2008), while the sixth trial included ballroom dancing (Haboush et al., 2006). All five resistance RCTs used progressive weight training (Favilla, 1992; Sims et al., 2006, 2009; Singh et al., 1997, 2005). Two mind-body RCTs were tai chi interventions (Cheng et al., 2012; Chou et al., 2004), whereas the other two were qigong interventions (Tsang et al., 2006; Tsang et al., 2013).

The average length (weeks) for exercise programs was longer for mind-body (M = 13.00, SD = 2.00), than aerobic (M = 9.33, SD = 4.32) or resistance (M = 8.29, SD = 1.80). The average minutes per week (calculated as frequency multiplied by duration) varied between exercise therapies, with the most time spent in resistance (M = 170.00, SD = 29.50), followed by mind-body (M = 140.63, SD = 28.31) and aerobic (M = 116.67, SD = 87.84).

Figure 2. Network plot of comparisons for all studies included in the network meta-analysis. Line width is proportional to the number of pairwise effect size estimates and node size is proportional to the number of participants.

3.3. Risk of Bias and Quality Assessment

Risk of bias within each study was generally low, with only three studies reporting substantial risk of bias across three of the seven criteria (Cheng et al., 2012; Chou et al., 2004; Lok et al., 2017). Full details on risk of bias for each study can be found in the Supplementary File S1. Notably, blinding of participants and personnel was not possible in exercise-based interventions, which resulted in a high risk of performance bias for all studies. The remaining six
risk of bias criteria were generally low-to-moderate across all studies (see Figure 3). Blinding of outcome assessment (detection bias) was not performed in three RCTs (Cheng et al., 2012; Chou et al., 2004; Lok et al., 2017), and incomplete outcome data (attrition bias) was identified in four trials (Favilla, 1992; Prakhinkit et al., 2014; Singh et al., 2005; Tsang et al., 2006). High ‘other sources of bias’ was generally due to small sample sizes (Cheng et al., 2012; Chou et al., 2004; McNeil et al., 1991; Patel, 2004; Sims et al., 2006), with one study reporting low adherence (Sims et al., 2009).

[Figure 3 Approximately Here]

Figure 3. Risk of bias chart of studies included in the quantitative analysis.

3.4. Secondary Outcomes

Secondary outcomes included adverse events, adherence, and attrition. No adverse events were reported in any of the included studies. Aerobic exercise had an adherence rate of 96%, which was based on statistics from a single study (Haboush et al., 2006). Six studies reported adherence rates for resistance exercise ($M = 93.0\%$, $SD = 9.27$, 95% CI = 83.27, 102.73). Mind-body exercise had an adherence rate of 95%, which was based on statistics from a single study (Chou et al., 2004). Attrition was higher for mind-body exercise (11.8%), followed by resistance exercise (10.5%), and aerobic exercise (7.6%).

3.5. Assessment of Inconsistency

Inconsistency network models were used to test the global consistency of direct and indirect effects for pairwise and multi-arm comparisons simultaneously. The assumption of consistency was satisfied for the overall level of each treatment ($p > .05$). Inconsistency tests
between direct and indirect estimates in the resistance versus attention-control comparison revealed that direct evidence was significantly different to indirect estimates (difference = -2.01; \( p < .05 \)). The remaining tests of inconsistency between direct and indirect estimates were non-significant (\( p > .05 \)), indicating that indirect estimates were not different to direct evidence.

Loop-specific heterogeneity was explored with an inconsistency plot. Inconsistency factors (\( IF \)) did not indicate high inconsistency (\( IF = 0.25 \) to 2.05), however loop-specific heterogeneity was detected in the AC-RES-WL loop (\( \tau^2 = 1.16 \)). No other loops reported heterogeneity or departed from the minimum lower-bound CI, and therefore, assumption of consistency was upheld. Full results from the inconsistency tests and plots can be found in the *Supplementary File S1*.

### 3.6. Publication Bias and Sensitivity Analyses

Publication bias and small-study effects were evaluated using a comparison-adjusted funnel plot (see *Figure 4*). The funnel plot was roughly symmetrical for the depression network, indicating no publication bias from small-study effects. Transitivity was explored using meta-regression sensitivity analyses. No significant modifying effects were found for year of study, risk of bias, publication status, intention-to-treat analysis, cluster design, age, gender, source of participants, length of intervention, format of program, exercise intensity, frequency, duration, or adherence, indicating that the assumption of transitivity was upheld. All meta-regression analyses can be found in the *Supplementary File S1*.

![Figure 4 Approximately Here](image)

*Figure 4.* Comparison-adjusted funnel plot for the depressive symptoms network. The red vertical line represents the null hypothesis that independent effect size estimates do not differ
from the comparison-specific pooled estimates. AC = attention-control; AER = aerobic; MB = mind-body; RES = resistance; UC = usual care; WL = wait-list.

3.7. Results of the Network Meta-Analysis

Data was pooled from the 15 studies and provided a total of 19 effect sizes. Network estimates were calculated to measure the relative effectiveness between pairs of comparisons within the network (see Figure 5 for full results of network meta-analysis). In brief, a greater decrease in depressive symptoms was observed for all exercise conditions when compared to control conditions. Indirect comparisons between the three exercise conditions demonstrated the strongest effect sizes for mind-body exercise, followed by aerobic exercise, and resistance exercise. Notably, the confidence intervals indicated that there is a large proportion of overlapping variance across these three indirect comparisons.

[Figure 5 Approximately Here]

Figure 5. Predictive interval plot for the depressive symptoms network. The black diamonds represent the difference in the effect size estimate (Hedges’ g). The narrow horizontal lines represent the confidence intervals (CI) and the wider horizontal lines represent the prediction intervals (PrI). The blue vertical line represents the null hypothesis (Hedges’ g = 0). Negative scores indicate a greater decrease in depressive symptom for the comparison (left) group. AC = attention-control; AER = aerobic; MB = mind-body; RES = resistance; UC = usual care; WL = wait-list.

3.8. GRADE Assessment
Certainty of evidence was assessed with the GRADE approach (Salanti et al., 2014), which can be seen in Table 2. We found moderate-to-low certainty in the findings from the aerobic networks, which was downgraded due to imprecision from small sample sizes and a risk of detection bias. Findings from the resistance networks had low certainty due to imprecision from small sample sizes, risk of detection or attrition bias, and inconsistency in the network estimates. The mind-body networks were high-to-moderate certainty, with downgrading mainly due to imprecision from small samples sizes and a risk of attrition bias. Certainty in the indirect comparisons between the three exercise therapies was low or very low, which was downgraded because imprecision in the confidence intervals indicated that the true effect could potentially favour either condition.

4. Discussion

In order of treatment effectiveness, mind-body exercise demonstrated the largest treatment effect on depressive symptoms compared with control conditions (i.e., wait-list, usual care, attention-control), followed closely by aerobic and resistance exercise, respectively. Risk of bias was generally low for the included RCTs, however small sample sizes and imprecision in effect size estimates limited the confidence in the indirect comparisons between aerobic, resistance, and mind-body exercise interventions.

4.1. Theoretical Implications for the Current Findings

The primary outcome of this network meta-analysis was depressive symptoms. In general, there was moderate confidence in the findings for aerobic and mind-body comparisons across all control conditions, while confidence in the resistance comparisons were downgraded to low certainty because of a high risk of bias and imprecision. Small study samples played a
major role in downgrading the certainty of the evidence. Furthermore, the dispersion in effect size estimates caused confidence intervals to become too wide to confidently conclude significant results, especially for the indirect comparisons between exercise interventions.

Nevertheless, the current findings indicated that aerobic exercise had a slightly greater effect than resistance exercise (Hedges’ $g = -0.10$, $PrI = -2.23, 2.03$), which is consistent with previous research (Penninx et al., 2002). Notably, the average length (weeks) for aerobic and resistance programs were comparable, whereas the average minutes spent per week (calculated as frequency multiplied by duration) was much higher in resistance conditions than aerobic conditions (170 minutes versus 117 minutes, respectively). This indicates that aerobic exercise can achieve a similar antidepressive treatment effect in a shorter duration than resistance exercise. Despite these trivial differences between aerobic and resistance exercise, both therapies appear to be effective in reducing clinical depression in older adulthood.

Five aerobic RCTs used walking (Lok et al., 2017; McNeil et al., 1991; Patel, 2004; Prakhinki et al., 2014; Williams & Tappen, 2008), while the remaining trial included ballroom dancing (Haboush et al., 2006). All five resistance RCTs used progressive weight training (Favilla, 1992; Sims et al., 2006, 2009; Singh et al., 1997, 2005), which is characterised by a gradual increase in the difficulty of exercises according to the individual’s own capabilities. Since these pools of studies reflect relatively homogeneous modalities of exercise, the findings indicate that clinically depressed older adults may respond similarly to aerobic walking and progressive resistance training.

Notable beneficial effects were observed for mind-body exercise, yet they did not reach the threshold for clinical significance when compared to aerobic and resistance exercise. In particular, interventions incorporating mind-body exercise therapy led to a relatively greater
reduction in depressive symptoms than those involving either aerobic (Hedges’ $g = -0.36$, $PrI = -2.69, 1.97$) or resistance exercise (Hedges’ $g = -0.46$, $PrI = -2.75, 1.83$), supporting previous meta-analytic subgroup analyses (Heinzel et al., 2015). Since mind-body exercise is typically performed at a lower intensity than other exercise modes, it appears that intensity may not be primarily responsible for the reduction in depression. Instead, physical exercise combined with an internally directed focus on breathing and proprioception may result in stronger antidepressive effects than physical exercise alone (La Forge, 1997; Prakhinkit et al., 2014). Hence, mind-body exercise can allow older adults to manage their depressive symptoms through interoceptive states not normally available during aerobic and resistance activities (Paulus & Stein, 2010).

Four mind-body trials were included in the current pool of studies, including two interventions using tai chi (Cheng et al., 2012; Chou et al., 2004) and two using qigong (Tsang et al., 2006, 2013). The length of these programs was evidently longer, with mind-body interventions averaging more than three weeks longer than aerobic and resistance interventions. Thus, it is plausible that the greater improvement in depressive symptomatology may be partially explained by participants experiencing a larger treatment dose.

Previously published clinical meta-analyses (Bridle et al., 2012; Heinzel et al., 2015; Schuch et al., 2016) included several studies which were not included in the current network meta-analysis. These studies were excluded because they did not fulfill our inclusion criteria, including a mean age lower than 65 years (Blumenthal et al., 2012; Mather et al., 2002) and incorporating a combination of exercise modalities in one condition (Brenes et al., 2007; Huang et al., 2015; Kerse et al., 2010). A further three studies were excluded because they used a multicomponent treatment that could have synergistically enhanced the treatment effect of exercise, such as medication (Belvederi Murri et al., 2015), depression management.
(Ciechanowski et al., 2004), or laughter therapy (Lavretsky et al., 2011). Our findings therefore provide a more accurate representation of the comparative effectiveness between aerobic, resistance, and mind-body exercise interventions on depressive symptoms for clinically depressed older adults.

4.2. Practical Implications

Secondary outcomes (i.e., adverse events, adherence, attrition) were used to evaluate overall compliance to exercise. No adverse events were identified in any studies, which means either (1) participants did not experience any major or minor adverse outcomes during exercise sessions, or (2) the researchers did not observe and/or record any adverse outcomes during the intervention period. Adherence was above 90 percent for all three exercise conditions, reflecting excellent adherence rates. Attrition rates were also compared between exercise interventions, which demonstrated slight discrepancies. Interestingly, highest dropout rates were observed for mind-body exercise, followed by resistance exercise. On the contrary, aerobic exercise had the lowest dropout rates, indicating that aerobic exercise is the best for high adherence to treatment while still maintaining a lower risk of attrition. Nevertheless, these outcomes demonstrate that older adults have very high compliance to physical exercise.

These findings reinforce the therapeutic benefit of exercise in clinically depressed older adults. Since non-compliance to antidepressant medication is a serious problem in older populations (Rossom et al., 2016; Stein-Shvachman et al., 2013), regular exercise can be incorporated into everyday life as a standalone treatment or an add-on strategy for other antidepressant therapies to further reduce depressive symptoms (Mura et al., 2014). This is particularly valuable for at-risk groups, such as those who are very old (Luppa et al., 2012), frail elderly (Soysal et al., 2017), people with dementia (De Souto Barreto et al., 2015), or residents in
aged care facilities (Seitz et al. 2010), who might also benefit from the physiological and social aspects of group exercise. Integrative treatments are essential to effectively treat many cases of geriatric depression. Therefore, it is recommended that allied health professionals and stakeholders consider the synergistic effects of physical exercise along with other factors such as diet, sleep patterns, social interaction, body-mindfulness, and lifestyle factors (Firth et al., 2019; Lopresti, 2019; Miller et al., 2019). It is also important that exercise programs are strategically developed to best suit the specific needs of older adults so that inequity can be addressed, allowing anyone to potentially benefit from a variety of exercise types and modalities. This can be achieved with an interdisciplinary approach to clinical treatment, as well as clinicians building a partnership with their patients to develop a modified exercise program to satisfy the physical needs, abilities, and personal interests of specific older individuals or groups.

4.3. Limitations and Future Directions

There are several noteworthy limitations in this network meta-analysis. Most importantly, the network meta-analysis was comprised of a relatively small group of participants ($n = 596$). Indeed, only two of the included RCTs had a sample size of over 30 participants for each comparison group (Lok et al., 2017; Tsang et al., 2006), introducing imprecision for the indirect comparisons. This is further exacerbated by the impossibility to blind participants and training personnel during exercise-based interventions, which introduced an inherent risk of performance bias. Risk of bias and GRADE assessments identified these factors as potential threats to the overall validity of the current findings. There was generally moderate-to-low confidence in the findings, and therefore, caution should be taken in the interpretation and implementation of the current findings.
It is also important to note that all exercise interventions from the included pool of RCTs lasted a maximum of 16 weeks, with an average length of 9.76 weeks ($SD = 3.38$ weeks). This is a rather short follow-up period, which is likely to reflect only the short-term impact of exercise on depressive symptoms. Despite this, exercise therapy seems to have a relatively strong impact on symptoms of clinical depression in older adults. The cumulative effect of long-term compliance to exercise therapy is still uncertain, and therefore, clinical applications are limited by the lack of available long-term research designs. This highlights the need for more high-quality clinical trials including intervention periods that are longer in duration.

Finally, treatment conditions using a combination of aerobic, resistance, and/or mind-body exercise were excluded during the screening process to prevent confounding effects of using more than one type of exercise within a single intervention group. While it is possible that a combination of exercise types may provide a synergistic or compound effect on depressive symptoms, this was beyond the scope of our review. Thus, it is recommended that future scientific reviews examine the effects of combined exercise modalities on depressive symptoms in clinically depressed older adults.

### 4.4. Conclusions

In summary, combined evidence from RCTs shows substantial and comparable beneficial treatment effects of aerobic, resistance, and mind-body exercise to reduce symptoms of clinical depression in older adults. This demonstrates that older adults can self-select their preferred exercise type in order to achieve optimal therapeutic benefit. In coalition with high levels of compliance, these findings should be used to advise allied health professionals and stakeholders in clinical geriatrics that clinically depressed older adults can benefit from the antidepressant effect of either aerobic, resistance, or mind-body exercise.
Acknowledgments

The authors graciously acknowledge the skilled contributions of Dr. Beyon Miloyan for his technical guidance, thoughtful discussions, and expert critique of the present work.

Appendix A Supplementary data

Supplementary material related to this article can be found in the online version at doi: https://doi.org/10.1016/j.arr.2019.100999.

Highlights

- Older adults can benefit from either aerobic, resistance, or mind-body exercise.
- Exercise is a therapeutic ally to pharmacological treatment of clinical depression.
- Pooled NMA evidence demonstrates high compliance and tolerance of exercise.
- There is opportunity for patients to select their preferred type(s) of exercise.
- Clinicians should facilitate exercise prescription based on patient preference.
References

*Studies included in the systematic review.


Mihalopoulos, C., Myles, H., O'Donoghue, B., Pillinger, T., Sarris, J., Schuch, F. B.,
Shiers, D., Smith, L., Solmi, M., Suetani, S., Taylor, J., Teasdale, S. B., Thorncroft, G.,
Torous, J., Usherwood, T., Vancampfort, D., Veronese, N., Ward, P. B., Yung, A. R.,
protecting physical health in people with mental illness. *The Lancet Psychiatry, 6*(8), 675-712. doi:10.1016/S2215-0366(19)30132-4


Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related


Higgins, J. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savović, J.,
assessing risk of bias in randomised trials. *British Medical Journal (Clinical Research
Ed.), 343*, d5928. doi:10.1136/bmj.d5928


Smarr, K. L., & Keefer, A. L. (2011). Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care & Research, 63*(Suppl. 11), S454-S466. doi:10.1002/acr.20556


StataCorp (2017). *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC.


### Table 1

<table>
<thead>
<tr>
<th>Author (year)</th>
<th>Treatment M&lt;sub&gt;age&lt;/sub&gt; (SD), females (%)</th>
<th>Control M&lt;sub&gt;age&lt;/sub&gt; (SD), females (%)</th>
<th>Source of participants</th>
<th>Inclusion criteria</th>
<th>Depression diagnosis</th>
<th>Intention-to-treat</th>
<th>Cluster design</th>
<th>Treatment group</th>
<th>Control group</th>
<th>Length of intervention</th>
<th>Adherence (%)</th>
<th>Outcome measure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cheng (2012)</td>
<td>81.0 (7.7) 50</td>
<td>82.5 (7.1) 75</td>
<td>Residential</td>
<td>Dementia</td>
<td>≥ 6 GDS-15 score</td>
<td>N/R</td>
<td>Yes (9)</td>
<td>n = 12</td>
<td>Mind-body (Tai Chi), supervised, group, N/R, 3 times/week, 60 mins/session</td>
<td>n = 12</td>
<td>Attention-control (handcrafts)</td>
<td>12 weeks</td>
</tr>
<tr>
<td>Chou (2004)</td>
<td>72.6 (4.2) 50</td>
<td></td>
<td>Community</td>
<td>Sedentary</td>
<td>≥ 16 CESD score; DSM-IV MDD or dysthymia diagnosis</td>
<td>N/R</td>
<td>No</td>
<td>n = 7</td>
<td>Mind-body (Tai Chi), supervised, group, N/R, 3 times/week, 45 mins/session</td>
<td>n = 7</td>
<td>Wait-list</td>
<td>12 weeks</td>
</tr>
<tr>
<td>Favilla (1992)</td>
<td>73 (4.0) 57.1</td>
<td>73 (4.0) 86.4</td>
<td>Community</td>
<td>Sedentary</td>
<td>≥ 7 GDS-30 score</td>
<td>N/R</td>
<td>No</td>
<td>n = 21</td>
<td>Resistance (progressive weight training), supervised, group, high intensity (80% 1R&lt;sub&gt;max&lt;/sub&gt;), 3 times/week, 60 mins/session</td>
<td>n = 22</td>
<td>Wait-list</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Haboush (2006)</td>
<td>69.4 (5.4) 67</td>
<td></td>
<td>Community</td>
<td>N/R</td>
<td>≥ 10 HRSD score</td>
<td>N/R</td>
<td>No</td>
<td>n = 10</td>
<td>Aerobic (ballroom dance), supervised, individual, N/R, 1 time/week, 45 mins/session</td>
<td>n = 12</td>
<td>Wait-list</td>
<td>8 weeks</td>
</tr>
<tr>
<td>Lok (2017)</td>
<td>65+ 42.5 65+ 47.5</td>
<td></td>
<td>Residential</td>
<td>N/R</td>
<td>≥ 10 BDI score</td>
<td>N/R</td>
<td>No</td>
<td>n = 40</td>
<td>Aerobic (walking and rhythmic exercises), supervised, group, N/R, 4 times/week, 70 mins/session</td>
<td>n = 40</td>
<td>Usual care</td>
<td>10 weeks</td>
</tr>
<tr>
<td>McNeil (1991)</td>
<td>72.5 (6.9) N/R</td>
<td></td>
<td>Community</td>
<td>N/R</td>
<td>12-24 BDI score</td>
<td>N/R</td>
<td>No</td>
<td>n = 10</td>
<td>Aerobic (walking), supervised, individual, N/R, 3 times/week, 20-40 mins/session</td>
<td>n = 10</td>
<td>Wait-list</td>
<td>6 weeks</td>
</tr>
<tr>
<td>Reference</td>
<td>Sample Size</td>
<td>Age (Mean ± SD)</td>
<td>Type</td>
<td>Score Threshold</td>
<td>Intervention Details</td>
<td>Sample Size</td>
<td>Duration</td>
<td>Control Group</td>
<td>Outcome Measures</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>----------------</td>
<td>------</td>
<td>----------------</td>
<td>----------------------</td>
<td>-------------</td>
<td>----------</td>
<td>---------------</td>
<td>------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patel (2004)</td>
<td>72.2 (N/R) 61.5 73.4 (N/R) 53.8</td>
<td>Residential N/R</td>
<td>≥ 5 GDS-15 score N/R No</td>
<td>n = 13</td>
<td>Aerobic (walking), supervised, group, N/R, 3 times/week, 20 mins/session</td>
<td>n = 13</td>
<td>4 weeks</td>
<td>N/R</td>
<td>GDS-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prakhinit (2014)</td>
<td>74.8 (1.7) 100 81.0 (1.7) 100</td>
<td>Community N/R</td>
<td>13-24 GDS-30 score N/R No</td>
<td>n = 13</td>
<td>Aerobic (walking), supervised, group, low intensity (20-50% HRmax), 3 times/week, 20-30 mins/session</td>
<td>n = 13</td>
<td>12 weeks</td>
<td>80+</td>
<td>GDS-30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shahidi (2011)</td>
<td>65.7 (4.2) 100 68.4 (6.3) 100</td>
<td>Community N/R</td>
<td>≥ 10 GDS-30 score N/R No</td>
<td>n = 20</td>
<td>Aerobic (jogging), supervised, group, N/R, N/R, 30 mins/session</td>
<td>n = 20</td>
<td>10 sessions</td>
<td>N/R</td>
<td>GDS-30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sims (2006)</td>
<td>75.3 (5.8) 85.7 74.3 (5.7) 50</td>
<td>Community N/R</td>
<td>≥ 11 GDS-30 score Yes No</td>
<td>n = 13</td>
<td>Resistance (progressive resistance training), supervised, individual, high intensity (80% 1Rmax), 3 times/week, 40 mins/session</td>
<td>n = 17</td>
<td>10 weeks</td>
<td>N/R</td>
<td>CESD-20*; GDS-30</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sims (2009)</td>
<td>68.0 (14.8) 39.1 66.3 (16.0) 40.9</td>
<td>Community History of stroke</td>
<td>≥ 5 PHQ-D score; psychiatrist depression confirmation Yes No</td>
<td>n = 13</td>
<td>Resistance (progressive resistance training), supervised, group, high intensity (80% 1Rmax), 2 times/week, N/R</td>
<td>n = 22</td>
<td>10 weeks</td>
<td>75</td>
<td>CESD-20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singh (1997)</td>
<td>70 (1.5) 70.6 72 (2.0) 53.3</td>
<td>Community Low activity (exercise no more than twice a week)</td>
<td>≥ 13 BDI score; DSM-IV MDD, minor depression, or dysthymia diagnosis N/R No</td>
<td>n = 17</td>
<td>Resistance (progressive resistance training), supervised, group, high intensity (80% 1Rmax), 3 times/week, 50 mins/session</td>
<td>n = 15</td>
<td>10 weeks</td>
<td>93</td>
<td>BDI*; GDS-30; HRSD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Singh (2005)</td>
<td>70 (7) 60 69 (7) 50</td>
<td>Community Low activity (exercise no more than twice a week)</td>
<td>≥ 14 GDS-30 score N/R No</td>
<td>n = 17</td>
<td>Resistance (progressive resistance training), supervised, group, low intensity (20% 1Rmax), 3 times/week, 65 mins/session</td>
<td>n = 19</td>
<td>8 weeks</td>
<td>N/R</td>
<td>GDS-30; HRSD*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>69 (5) 55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Year</td>
<td>Control Group</td>
<td>Intervention Group</td>
<td>No.</td>
<td>Intervention Details</td>
<td>Duration</td>
<td>Follow-up</td>
<td>Outcome Measure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>---------------</td>
<td>--------------------</td>
<td>-----</td>
<td>--------------------------------------------------------------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tsang (2006)</td>
<td>82.1(7.2) 79.2</td>
<td>Residential</td>
<td>N/R</td>
<td>≥ 9 GDS-15 score</td>
<td>N/R</td>
<td>No</td>
<td>n = 48</td>
<td>Mind-body (Baduanjin qigong), supervised, group, N/R, 3 times/week, 30-45 mins/session</td>
<td>16 weeks</td>
<td>N/R</td>
<td>GDS-15</td>
<td></td>
</tr>
<tr>
<td>Tsang (2013)</td>
<td>79.2 (6.6) 80.7 (4.4)</td>
<td>Residential</td>
<td>Chronic medical illness</td>
<td>Yes</td>
<td>≥ 8 GDS-15 score; DSM-IV MDD diagnosis</td>
<td>N/R</td>
<td>No</td>
<td>n = 21</td>
<td>Mind-body (eight-section brocades qigong), supervised, group, N/R, 3 times/week, 45 mins/session</td>
<td>12 weeks</td>
<td>N/R</td>
<td>GDS-15; HRSD*</td>
</tr>
<tr>
<td>Williams (2008)</td>
<td>87.9 (6.0) 88.9</td>
<td>Residential</td>
<td>Low activity (walked unaided for 30 minutes or more); Dementia</td>
<td>≥ 7 CSDD score</td>
<td>Yes</td>
<td>No</td>
<td>n = 17</td>
<td>Aerobic (walking), supervised, individual, N/R, 5 times/week, 30 mins/session</td>
<td>16 weeks</td>
<td>N/R</td>
<td>CSDD</td>
<td></td>
</tr>
</tbody>
</table>

Note. N/R = not reported; n = Total participants included in intention-to-treat or post-treatment data analysis; MDD = Major depressive disorder; DSM-IV = Diagnostic and Statistical Manual of Mental Disorders (4th Edition); PHQ-D = Patient Health Questionnaire (depression subscale); HRmax = Heart rate maximum; 1Rmax = One-repetition maximum; RPE = Rating of perceived exertion; BDI = Beck Depression Inventory; CESD = Center for Epidemiological Studies Depression Scale; CSDD = Cornell Scale for Depression in Dementia; GDS = Geriatric Depression Scale; HRSD = Hamilton Rating Scale for Depression.

*Outcome measure used in network meta-analysis.
# Table 2

**Summary of GRADE assessment for the certainty in depressive symptoms estimates**

<table>
<thead>
<tr>
<th>Comparison Effect</th>
<th>Number of Participants</th>
<th>Number of Direct Comparisons</th>
<th>Nature of Evidence</th>
<th>Certainty</th>
<th>Reason for Downgrading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aerobic vs. wait-list</td>
<td>20 vs. 22</td>
<td>2</td>
<td>Mixed</td>
<td>Moderate</td>
<td>Imprecision&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aerobic vs. usual care</td>
<td>66 vs. 66</td>
<td>3</td>
<td>Mixed</td>
<td>Moderate</td>
<td>Risk of bias&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aerobic vs. attention-control</td>
<td>27 vs. 22</td>
<td>2</td>
<td>Mixed</td>
<td>Low</td>
<td>Imprecision&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Resistance vs. wait-list</td>
<td>65 vs. 44</td>
<td>3</td>
<td>Mixed</td>
<td>Low</td>
<td>Risk of bias&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Resistance vs. usual care</td>
<td>35 vs. 19</td>
<td>2</td>
<td>Mixed</td>
<td>Low</td>
<td>Imprecision&lt;sup&gt;e&lt;/sup&gt;, risk of bias&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Resistance vs. attention-control</td>
<td>30 vs. 32</td>
<td>2</td>
<td>Mixed</td>
<td>Very low</td>
<td>Imprecision&lt;sup&gt;f&lt;/sup&gt;, inconsistency&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mind-body vs. wait-list</td>
<td>7 vs. 7</td>
<td>1</td>
<td>Mixed</td>
<td>Moderate</td>
<td>Imprecision&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mind-body vs. usual care</td>
<td>0 vs. 0</td>
<td>0</td>
<td>Indirect</td>
<td>Moderate</td>
<td>Imprecision&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mind-body vs. attention-control</td>
<td>81 vs. 63</td>
<td>3</td>
<td>Mixed</td>
<td>Moderate</td>
<td>Risk of bias&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aerobic vs. resistance</td>
<td>0 vs. 0</td>
<td>0</td>
<td>Indirect</td>
<td>Low</td>
<td>Imprecision&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Aerobic vs. mind-body</td>
<td>0 vs. 0</td>
<td>0</td>
<td>Indirect</td>
<td>Low</td>
<td>Imprecision&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Resistance vs. mind-body</td>
<td>0 vs. 0</td>
<td>0</td>
<td>Indirect</td>
<td>Low</td>
<td>Imprecision&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>Small sample size.

<sup>b</sup>Potential detection bias due to high number of studies without blinding of outcome assessment.

<sup>c</sup>Wide confidence intervals favouring either treatment.

<sup>d</sup>Potential attrition bias due to high number of studies with incomplete outcome data.

<sup>e</sup>Evidence of inconsistency in the network.
Figure 1

Records identified through database searching
- PubMed (n = 892)
- Web of Science (n = 1,654)
- CINAHL (n = 515)
- Health Source: N&H, Ed. (n = 153)
- PsycARTICLES (n = 10)
- PsycINFO (n = 349)
- SPORTDiscus (n = 85)

Records identified through other sources (n = 26)

Total records identified (n = 3,704)

Duplicate records removed (n = 1,395)

Records screened by titles and abstracts (n = 2,309)

Records not relevant to primary outcome excluded (n = 1,953)

Full-text articles excluded (n = 340)
- Age below 65 years (n = 76)
- Not clinically depressed at baseline (n = 81)
- Not randomly assigned (n = 21)
- No control group (n = 39)
- No exercise group (n = 16)
- Depression not measured at follow-up (n = 24)
- Mixed exercise condition (n = 47)
- Duplicate sample set (n = 12)
- Full-text unobtainable (n = 1)

Full-text articles assessed for eligibility (n = 356)

Studies included in systematic review (n = 16)

No clearly defined control group (n = 1)

Studies included in quantitative analysis (n = 15)
Figure 3

[Diagram showing various sources of bias with color-coded bars indicating low, unclear, and high levels]
Figure 4
Figure 5

Comparison Effect

<table>
<thead>
<tr>
<th></th>
<th>AC vs WL</th>
<th>AER vs AC</th>
<th>MB vs AER</th>
<th>RES vs MB</th>
<th>UC vs RES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-0.28 (-1.23, 0.67)</td>
<td>-0.51 (-1.46, 0.45)</td>
<td>-0.36 (-1.59, 0.87)</td>
<td>0.46 (-0.72, 1.63)</td>
<td>0.92 (-0.33, 1.87)</td>
</tr>
<tr>
<td>95% CI</td>
<td>(-2.41, 1.65)</td>
<td>(-2.64, 1.63)</td>
<td>(-2.69, 1.97)</td>
<td>(-1.83, 2.75)</td>
<td>(-1.21, 3.05)</td>
</tr>
<tr>
<td>95% PrI</td>
<td>(-3.43, 1.13)</td>
<td>(-2.85, 1.22)</td>
<td>(-2.51, 1.89)</td>
<td>(0.04, 2.72)</td>
<td>(-1.04, 3.80)</td>
</tr>
</tbody>
</table>

Favours First Comparison  Favours Second Comparison