

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Noetel, Michael; Sanders, Taren; Gallardo-Gómez, Daniel; Taylor, Paul; del Pozo Cruz, Borja; van den Hoek, Daniel; Smith, Jordan J; Mahoney, John; Spathis, Jemima; Moresi, Mark; Pagano, Rebecca; Pagano, Lisa; Vasconcellos, Roberta; Arnott, Hugh; Varley, Benjamin; Parker, Philip; Biddle, Stuart; Lonsdale, Chris

Title: Effect of exercise for depression : systematic review and network meta-analysis of randomised controlled trials

Year: 2024

Version: Published version

Copyright: © Authors 2024

Rights: CC BY-NC 4.0

Rights url: <https://creativecommons.org/licenses/by-nc/4.0/>

Please cite the original version:

Noetel, M., Sanders, T., Gallardo-Gómez, D., Taylor, P., del Pozo Cruz, B., van den Hoek, D., Smith, J. J., Mahoney, J., Spathis, J., Moresi, M., Pagano, R., Pagano, L., Vasconcellos, R., Arnott, H., Varley, B., Parker, P., Biddle, S., & Lonsdale, C. (2024). Effect of exercise for depression : systematic review and network meta-analysis of randomised controlled trials. *BMJ*, 384, Article e075847. <https://doi.org/10.1136/bmj-2023-075847>



OPEN ACCESS



Check for updates

Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials

Michael Noetel,¹ Taren Sanders,² Daniel Gallardo-Gómez,³ Paul Taylor,⁴ Borja del Pozo Cruz,^{5,6} Daniel van den Hoek,⁷ Jordan J Smith,⁸ John Mahoney,⁹ Jemima Spathis,⁹ Mark Moresi,⁴ Rebecca Pagano,¹⁰ Lisa Pagano,¹¹ Roberta Vasconcellos,² Hugh Arnott,² Benjamin Varley,¹² Philip Parker,¹³ Stuart Biddle,^{14,15} Chris Lonsdale¹³

For numbered affiliations see end of the article

Correspondence to: M Noetel
m.noetel@uq.edu.au
(or @mnoetel on Twitter;
ORCID 0000-0002-6563-8203)
Additional material is published
online only. To view please visit
the journal online.

Cite this as: *BMJ* 2024;384:e075847
<http://dx.doi.org/10.1136/bmj-2023-075847>

Accepted: 15 January 2024

ABSTRACT OBJECTIVE

To identify the optimal dose and modality of exercise for treating major depressive disorder, compared with psychotherapy, antidepressants, and control conditions.

DESIGN

Systematic review and network meta-analysis.

METHODS

Screening, data extraction, coding, and risk of bias assessment were performed independently and in duplicate. Bayesian arm based, multilevel network meta-analyses were performed for the primary analyses. Quality of the evidence for each arm was graded using the confidence in network meta-analysis (CINeMA) online tool.

DATA SOURCES

Cochrane Library, Medline, Embase, SPORTDiscus, and PsycINFO databases.

ELIGIBILITY CRITERIA FOR SELECTING STUDIES

Any randomised trial with exercise arms for participants meeting clinical cut-offs for major depression.

RESULTS

218 unique studies with a total of 495 arms and 14 170 participants were included. Compared with active controls (eg, usual care, placebo tablet), moderate reductions in depression were found for walking or jogging (n=1210, $\kappa=51$, Hedges' g -0.62 , 95% credible interval -0.80 to -0.45), yoga (n=1047, $\kappa=33$, g -0.55 , -0.73 to -0.36), strength training (n=643, $\kappa=22$, g -0.49 , -0.69 to -0.29), mixed aerobic exercises (n=1286, $\kappa=51$, g -0.43 , -0.61 to -0.24), and tai chi or qigong (n=343, $\kappa=12$,

g -0.42 , -0.65 to -0.21). The effects of exercise were proportional to the intensity prescribed. Strength training and yoga appeared to be the most acceptable modalities. Results appeared robust to publication bias, but only one study met the Cochrane criteria for low risk of bias. As a result, confidence in accordance with CINeMA was low for walking or jogging and very low for other treatments.

CONCLUSIONS

Exercise is an effective treatment for depression, with walking or jogging, yoga, and strength training more effective than other exercises, particularly when intense. Yoga and strength training were well tolerated compared with other treatments. Exercise appeared equally effective for people with and without comorbidities and with different baseline levels of depression. To mitigate expectancy effects, future studies could aim to blind participants and staff. These forms of exercise could be considered alongside psychotherapy and antidepressants as core treatments for depression.

SYSTEMATIC REVIEW REGISTRATION

PROSPERO CRD42018118040.

Introduction

Major depressive disorder is a leading cause of disability worldwide¹ and has been found to lower life satisfaction more than debt, divorce, and diabetes² and to exacerbate comorbidities, including heart disease,³ anxiety,⁴ and cancer.⁵ Although people with major depressive disorder often respond well to drug treatments and psychotherapy,^{6,7} many are resistant to treatment.⁸ In addition, access to treatment for many people with depression is limited, with only 51% treatment coverage for high income countries and 20% for low and lower-middle income countries.⁹ More evidence based treatments are therefore needed.

Exercise may be an effective complement or alternative to drugs and psychotherapy.¹⁰⁻¹⁴ In addition to mental health benefits, exercise also improves a range of physical and cognitive outcomes.¹⁵⁻¹⁷ Clinical practice guidelines in the US, UK, and Australia recommend physical activity as part of treatment for depression.¹⁸⁻²¹ But these guidelines do not provide clear, consistent recommendations about dose or exercise modality. British guidelines recommend group exercise programmes^{20,21} and offer general recommendations to increase any form of physical activity,²¹ the American Psychiatric Association recommends any dose of aerobic exercise or resistance training,²⁰ and Australian and New Zealand guidelines

WHAT IS ALREADY KNOWN ON THIS TOPIC

Depression is a leading cause of disability, and exercise is often recommended alongside first line treatments such as pharmacotherapy and psychotherapy. Treatment guidelines and previous reviews disagree on how to prescribe exercise to best treat depression.

WHAT THIS STUDY ADDS

Various exercise modalities are effective (walking, jogging, mixed aerobic exercise, strength training, yoga, tai chi, qigong) and well tolerated (especially strength training and yoga).

Effects appeared proportional to the intensity of exercise prescribed and were stronger for group exercise and interventions with clear prescriptions.

Preliminary evidence suggests interactions between types of exercise and patients' personal characteristics.



Effect of exercise for depression

A systematic review and network meta-analysis of randomised controlled trials

Summary



For treating depression, various exercise modalities are well tolerated and effective, particularly walking or jogging, yoga, and strength training. Effects were comparable to psychotherapy and pharmacotherapy. Exercise worked better when more intense

Population

14 170 participants

Participants with depression (ie, meeting clinical thresholds or diagnosed by a clinician)

No other exclusion criteria, so participants are from any age, and with any comorbidities

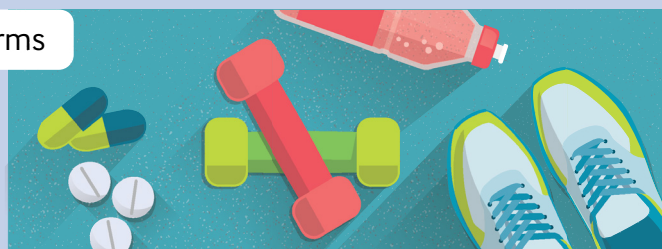
Study design

218 studies

495 unique arms

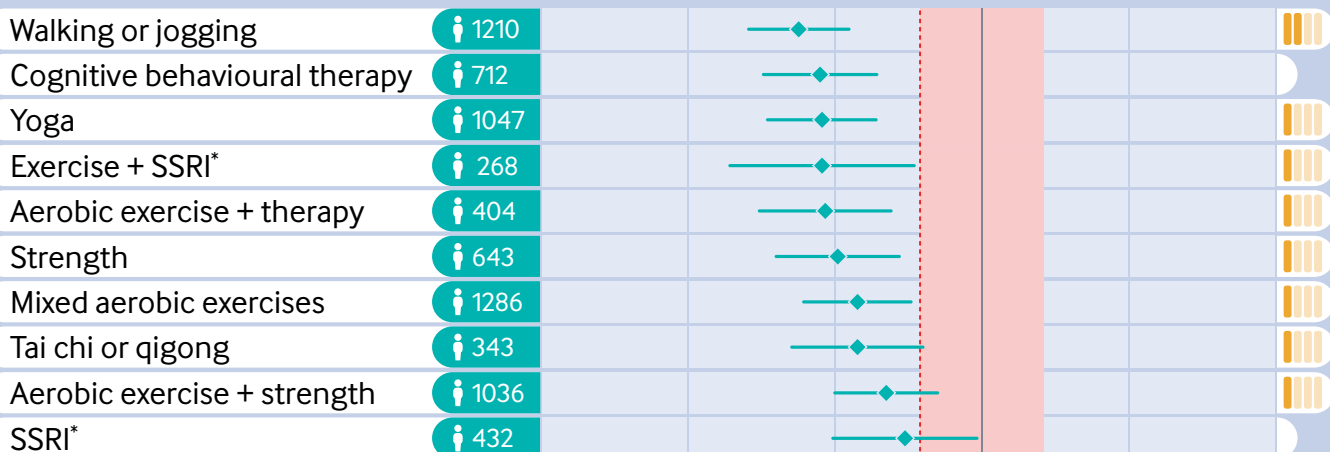
Comparison

Different forms of exercise compared with other common treatments for depression. All results are presented as 'compared with active controls.'



Outcomes

Network meta-analysis, standardised mean difference 95% CrI



Certainty rating

Low
Very low

--- Clinically important benefit

Equivalent to active control

* Selective serotonin reuptake inhibitor

<https://bit.ly/BMJ-exedep>

© 2024 BMJ Publishing Group Ltd

suggest a combination of strength and vigorous aerobic exercises, with at least two or three bouts weekly.¹⁹

Authors of guidelines may find it hard to provide consistent recommendations on the basis of existing mainly pairwise meta-analyses—that is, assessing a specific modality versus a specific comparator in a distinct group of participants.^{12 13 22} These meta-analyses have come under scrutiny for pooling heterogeneous treatments and heterogeneous comparisons leading to ambiguous effect estimates.²³ Reviews also face the opposite problem, excluding exercise treatments such as yoga, tai chi, and qigong

because grouping them with strength training might be inappropriate.²³ Overviews of reviews have tried to deal with this problem by combining pairwise meta-analyses on individual treatments. A recent such overview found no differences between exercise modalities.¹³ Comparing effect sizes between different pairwise meta-analyses can also lead to confusion because of differences in analytical methods used between meta-analysis, such as choice of a control to use as the referent. Network meta-analyses are a better way to precisely quantify differences between interventions as they simultaneously

model the direct and indirect comparisons between interventions.²⁴

Network meta-analyses have been used to compare different types of psychotherapy and pharmacotherapy for depression.^{6 25 26} For exercise, they have shown that dose and modality influence outcomes for cognition,¹⁶ back pain,¹⁵ and blood pressure.¹⁷ Two network meta-analyses explored the effects of exercise on depression: one among older adults²⁷ and the other for mental health conditions.²⁸ Because of the inclusion criteria and search strategies used, these reviews might have been under-powered to explore moderators such as dose and modality ($\kappa=15$ and $\kappa=71$, respectively). To resolve conflicting findings in existing reviews, we comprehensively searched randomised trials on exercise for depression to ensure our review was adequately powered to identify the optimal dose and modality of exercise. For example, a large overview of reviews found effects on depression to be proportional to intensity, with vigorous exercise appearing to be better,¹³ but a later meta-analysis found no such effects.²² We explored whether recommendations differ based on participants' sex, age, and baseline level of depression.

Given the challenges presented by behaviour change in people with depression,²⁹ we also identified autonomy support or behaviour change techniques that might improve the effects of intervention.³⁰ Behaviour change techniques such as self-monitoring and action planning have been shown to influence the effects of physical activity interventions in adults (>18 years)³¹ and older adults (>60 years)³² with differing effectiveness of techniques in different populations. We therefore tested whether any intervention components from the behaviour change technique taxonomy were associated with higher or lower intervention effects.³⁰ Other meta-analyses found that physical activity interventions work better when they provide people with autonomy (eg, choices, invitational language).³³ Autonomy is not well captured in the taxonomy for behaviour change technique. We therefore tested whether effects were stronger in studies that provided more autonomy support to patients. Finally, to understand the mechanism of intervention effects, such as self-confidence, affect, and physical fitness, we collated all studies that conducted formal mediation analyses.

Methods

Our findings are presented according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Network Meta-analyses (PRISMA-NMA) guidelines (see supplementary file, section S0; all supplementary files, data, and code are also available at <https://osf.io/nzw6u/>).³⁴ We amended our analysis strategy after registering our review; these changes were to better align with new norms established by the Cochrane Comparing Multiple Interventions Methods Group.³⁵ These norms were introduced between the publication of our protocol and the preparation of this manuscript. The largest change

was using the confidence in network meta-analysis (CINeMA)³⁵ online tool instead of the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) guidelines and adopting methods to facilitate assessments—for example, instead of using an omnibus test for all treatments, we assessed publication bias for each treatment compared with active controls. We also modelled acceptability (through dropout rate), which was not predefined but was adopted in response to a reviewer's comment.

Eligibility criteria

To be eligible for inclusion, studies had to be randomised controlled trials that included exercise as a treatment for depression and included participants who met the criteria for major depressive disorder, either clinician diagnosed or identified through participant self-report as exceeding established clinical thresholds (eg, scored >13 on the Beck depression inventory-II).³⁶ Studies could meet these criteria when all the participants had depression or when the study reported depression outcomes for a subgroup of participants with depression at the start of the study.

We defined exercise as “planned, structured and repetitive bodily movement done to improve or maintain one or more components of physical fitness.”³⁷ Unlike recent reviews,^{12 22} we included studies with more than one exercise arm and multifaceted interventions (eg, health and exercise counselling) as long as they contained a substantial exercise component. These trials could be included because network meta-analysis methods allows for the grouping of those interventions into homogenous nodes. Unlike the most recent Cochrane review,¹² we also included participants with physical comorbidities such as arthritis and participants with postpartum depression because the *Diagnostic Statistical Manual of Mental Health Disorders*, fifth edition, removed the postpartum onset specifier after that analysis was completed.²³ Studies were excluded if interventions were shorter than one week, depression was not reported as an outcome, and data were insufficient to calculate an effect size for each arm. Any comparison condition was included, allowing us to quantify the effects against established treatments (eg, selective serotonin reuptake inhibitors (SSRIs), cognitive behavioural therapy), active control conditions (usual care, placebo tablet, stretching, educational control, and social support), or waitlist control conditions. Published and unpublished studies were included, with no restrictions on language applied.

Information sources

We adapted the search strategy from the most recent Cochrane review,¹² adding keywords for yoga, tai chi, and qigong, as they met our definition for exercise. We conducted database searches, without filters or date limits, in The Cochrane Library via CENTRAL, SPORTDiscus via Embase, and Medline, Embase, and PsycINFO via Ovid. Searches of the databases were conducted on 17 December 2018 and 7 August 2020

and last updated on 3 June 2023 (see supplementary file section S1 for full search strategies). We assessed full texts of all included studies from two systematic reviews of exercise for depression.^{12 22}

Study selection and data collection

To select studies, we removed duplicate records in Covidence³⁸ and then screened each title and abstract independently and in duplicate. Conflicts were resolved through discussion or consultation with a third reviewer. The same methods were used for full text screening.

We used the Extraction 1.0 randomised controlled trial data extraction forms in Covidence.³⁸ Data were extracted independently and in duplicate, with conflicts resolved through discussion with a third reviewer.

Data items

For each study, we extracted a description of the interventions, including frequency, intensity, and type and time of each exercise intervention. Using the Compendium of Physical Activities,³⁹ we calculated the energy expenditure dose of exercise for each arm as metabolic equivalents of task (METs) min/week. Two authors evaluated each exercise intervention using the Behaviour Change Taxonomy version 1³⁰ for behaviour change techniques explicitly described in each exercise arm. They also rated the level of autonomy offered to participants, on a scale from 1 (no choice) to 10 (full autonomy). We also extracted descriptions of the other arms within the randomised trials, including other treatment or control conditions; participants' age, sex, comorbidities, and baseline severity of depressive symptoms; and each trial's location and whether or not the trial was funded.

Risk of bias in individual studies

We used Cochrane's risk of bias tool for randomised controlled trials.⁴⁰ Risk of bias was rated independently and in duplicate, with conflicts resolved through discussion with a third reviewer.

Summary measures and synthesis

For main and moderation analyses, we used bayesian arm based multilevel network meta-analysis models.⁴¹ All network meta-analytical approaches allow users to assess the effects of treatments against a range of comparisons. The bayesian arm based models allowed us to also assess the influence of hypothesised moderators, such as intensity, dose, age, and sex. Many network meta-analyses use contrast based methods, comparing post-test scores between study arms.⁴¹ Arm based meta-analyses instead describe the population-averaged absolute effect size for each treatment arm (ie, each arm's change score).⁴¹ As a result, the summary measure we used was the standardised mean change from baseline, calculated as standardised mean differences with correction for small studies (Hedges' g). In keeping with the norms from the included studies, effect sizes describe treatment effects

on depression, such that larger negative numbers represent stronger effects on symptoms. Using National Institute for Health and Care Excellence guidelines,⁴² we standardised change scores for different depression scales (eg, Beck depression inventory, Hamilton depression rating scale) using an internal reference standard for each scale (for each scale, the average of pooled standard deviations at baseline) reported in our meta-analysis. Because depression scores generally show regression to the mean, even in control conditions, we present effect sizes as improvements beyond active control conditions. This convention makes our results comparable to existing, contrast based meta-analyses.

Active control conditions (usual care, placebo tablet, stretching, educational control, and social support) were grouped to increase power for moderation analyses, for parsimony in the network graph, and because they all showed similar arm based pooled effect sizes (Hedges' g between -0.93 and -1.00 for all, with no statistically significant differences). We separated waitlist control from these active control conditions because it typically shows poorer effects in treatment for depression.⁴³

Bayesian meta-analyses were conducted in R⁴⁴ using the *brms* package.⁴⁵ We preregistered informative priors based on the distributional parameters of our meta-analytical model.⁴⁶ We nested effects within arms to manage dependency between multiple effect sizes from the same participants.⁴⁶ For example, if one study reported two self-reported measures of depression, or reported both self-report and clinician rated depression, we nested these effect sizes within the arm to account for both pieces of information while controlling for dependency between effects.⁴⁶ Finally, we compared absolute effect sizes against a standardised minimum clinically important difference, 0.5 standard deviations of the change score.⁴⁷ From our data, this corresponded to a large change in before and after scores (Hedges' g -1.16), a moderate change compared with waitlist control (g -0.55), or a small benefit when compared with active controls (g -0.20). For credibility assessments comparing exercise modalities, we used the *netmeta* package⁴⁸ and CINeMA.⁴⁹ We also used *netmeta* to model acceptability, comparing the odds ratio for drop-out rate in each arm.

Additional analyses

All prespecified moderation and sensitivity analyses were performed. We moderated for participant characteristics, including participants' sex, age, baseline symptom severity, and presence or absence of comorbidities; duration of the intervention (weeks); weekly dose of the intervention; duration between completion of treatment and measurement, to test robustness to remission (in response to a reviewer's suggestion); amount of autonomy provided in the exercise prescription; and presence of each behaviour change technique. As preregistered, we moderated for behaviour change techniques in three ways: through meta-regression, including all behaviour change

techniques simultaneously for primary analysis; including one behaviour change technique at a time (using 99% credible intervals to somewhat control for multiple comparisons) in exploratory analyses; and through meta-analytical classification and regression trees (metaCART), which allowed for interactions between moderating variables (eg, if goal setting combined with feedback had synergistic effects).⁵⁰ We conducted sensitivity analyses for risk of bias, assessing whether studies with low versus unclear or high risk of bias on each domain showed statistically significant differences in effect sizes.

Credibility assessment

To assess the credibility of each comparison against active control, we used CINeMA.^{35 49} This online tool was designed by the Cochrane Comparing Multiple Interventions Methods Group as an adaptation of GRADE for network meta-analyses.³⁵ In line with recommended guidelines, for each comparison we made judgements for within study bias, reporting bias, indirectness, imprecision, heterogeneity, and incoherence. Similar to GRADE, we considered the evidence for comparisons to show high confidence then downgraded on the basis of concerns in each domain, as follows:

Within study bias—Comparisons were downgraded when most of the studies providing direct evidence for comparisons were unclear or high risk.

Reporting bias—Publication bias was assessed in three ways. For each comparison with at least 10 studies⁵¹ we created funnel plots, including estimates of effect sizes after removing studies with statistically significant findings (ie, worst case estimates)⁵²; calculated an *s* value, representing how strong publication bias would need to be to nullify meta-analytical effects⁵²; and conducted a multilevel Egger's regression test, indicative of small study bias. Given these tests are not recommended for comparisons with fewer than 10 studies,⁵¹ those comparisons were considered to show "some concerns."

Indirectness—Our primary population of interest was adults with major depression. Studies were considered to be indirect if they focused on one sex only (>90% male or female), participants with comorbidities (eg, heart disease), adolescents and young adults (14-20 years), or older adults (>60 years). We flagged these studies as showing some concerns if one of these factors was present, and as "major concerns" if two of these factors were present. Evidence from comparisons was classified as some concerns or major concerns using majority rating for studies directly informing the comparison.

Imprecision—As per CINeMA, we used the clinically important difference of Hedges' $g=0.2$ to ascribe a zone of equivalence, where differences were not considered clinically significant ($-0.2 < g < 0.2$). Studies were flagged as some concerns for imprecision if the bounds of the 95% credible interval extended across that zone, and they were flagged as major concerns if the bounds extended to the other side of the zone of equivalence (such that effects could be harmful).

Heterogeneity—Prediction intervals account for heterogeneity differently from credible intervals.³⁵ As a result, CINeMA accounts for heterogeneity by assessing whether the prediction intervals and the credible intervals lead to different conclusions about clinical significance (using the same zone of equivalence from imprecision). Comparisons are flagged as some concerns if the prediction interval crosses into, or out of, the zone of equivalence once (eg, from helpful to no meaningful effect), and as major concerns if the prediction interval crosses the zone twice (eg, from helpful and harmful).

Incoherence—Incoherence assesses whether the network meta-analysis provides similar estimates when using direct evidence (eg, randomised controlled trials on strength training versus SSRI) compared with indirect evidence (eg, randomised controlled trials where either strength training or SSRI uses waitlist control). Incoherence provides some evidence the network may violate the assumption of transitivity: that the only systematic difference between arms is the treatment, not other confounders. We assessed incoherence using two methods: Firstly, a global design-by-treatment interaction to assess for incoherence across the whole network,^{35 49} and, secondly, separating indirect and direct evidence (SIDE method) for each comparison through netsplitting to see whether differences between those effect estimates were statistically significant. We flagged comparisons as some concerns if either no direct comparisons were available or direct and indirect evidence gave different conclusions about clinical significance (eg, from helpful to no meaningful effect, as per imprecision and heterogeneity). Again, we classified comparisons as major concerns if the direct and indirect evidence changed the sign of the effect or changed both limits of the credible interval.^{35 49}

Patient and public involvement

We discussed the aims and design of this study with members of the public, including those who had experienced depression. Several of our authors have experienced major depressive episodes, but beyond that we did not include patients in the conduct of this review.

Results

Study selection

The PRISMA flow diagram outlines the study selection process (fig 1). We used two previous reviews to identify potentially eligible studies for inclusion.^{12 22} Database searches identified 18 658 possible studies. After 5505 duplicates had been removed, two reviewers independently screened 13 115 titles and abstracts. After screening, two reviewers independently reviewed 1738 full text articles. Supplementary file section S2 shows the consensus reasons for exclusion. A total of 218 unique studies described in 246 reports were included, totalling 495 arms and 14 170 participants. Supplementary file section S3 lists the references and characteristics of the included studies.

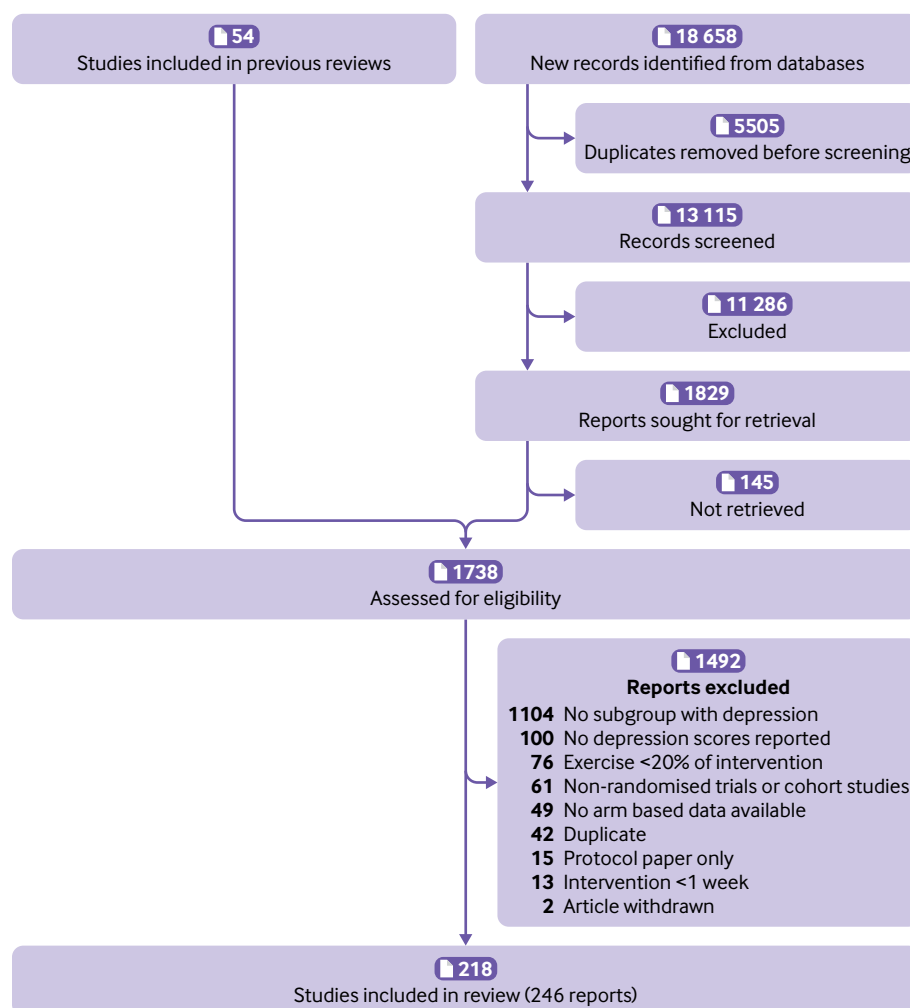


Fig 1 | Flow of studies through review

Network geometry

As preregistered, we removed nodes with fewer than 100 participants. Using this filter, most interventions contained comparisons with at least four other nodes in the network geometry (fig 2). The results of the global test design-by-treatment interaction model were not statistically significant, supporting the assumption of transitivity ($\chi^2=94.92$, $df=75$, $P=0.06$). When net-splitting was used on all possible combinations in the network, for two out of the 120 comparisons we found statistically significant incoherence between direct and indirect evidence (SSRI v waitlist control; cognitive behavioural therapy v tai chi or qigong). Overall, we found little statistical evidence that the model violated the assumption of transitivity. Qualitative differences were, however, found for participant characteristics between different arms (see supplementary file, section S4). For example, some interventions appeared to be prescribed more frequently among people with severe depression (eg, 7/16 studies using SSRIs) compared with other interventions (eg, 1/15 studies using aerobic exercise combined with therapy). Similarly, some interventions appeared more likely to be prescribed for older adults (eg, mean age, tai chi=59 v

dance=31) or women (eg, per cent female: dance=88% v cycling=53%). Given that plausible mechanisms exist for these systematic differences (eg, the popularity of tai chi among older adults),⁵³ there are reasons to believe that allocation to treatment arms would be less than perfectly random. We have factored these biases in our certainty estimates through indirectness ratings.

Risk of bias within studies

Supplementary file section S5 provides the risk of bias ratings for each study. Few studies explicitly blinded participants and staff (fig 3). As a result, overall risk of bias for most studies was unclear or high, and effect sizes could include expectancy effects, among other biases. However, sensitivity analyses suggested that effect sizes were not influenced by any risk of bias criteria owing to wide credible intervals (see supplementary file, section S6). Nevertheless, certainty ratings for all treatment arms were downgraded owing to high risk of bias in the studies informing the comparison.

Synthesis of results

Supplementary file section S7 presents a forest plot of Hedges' g values for each study. Figure 4 shows the

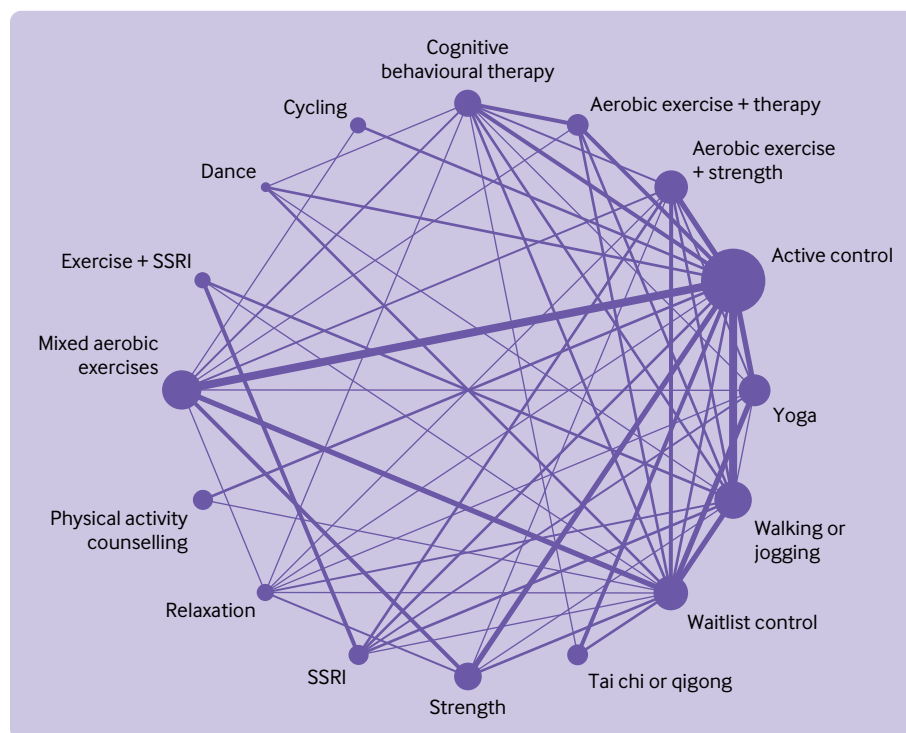


Fig 2 | Network geometry indicating number of participants in each arm (size of points) and number of comparisons between arms (thickness of lines). SSRI=selective serotonin reuptake inhibitor

predicted effects of each treatment compared with active controls. Compared with active controls, large reductions in depression were found for dance ($n=107$, $\kappa=5$, Hedges' $g = -0.96$, 95% credible interval -1.36 to -0.56) and moderate reductions for walking or jogging ($n=1210$, $\kappa=51$, $g = -0.63$, -0.80 to -0.46), yoga ($n=1047$, $\kappa=33$, $g = -0.55$, -0.73 to -0.36), strength training ($n=643$, $\kappa=22$, $g = -0.49$, -0.69 to -0.29), mixed aerobic exercises ($n=1286$, $\kappa=51$, $g = -0.43$, -0.61 to -0.25), and tai chi or qigong ($n=343$, $\kappa=12$,

$g = -0.42$, -0.65 to -0.21). Moderate, clinically meaningful effects were also present when exercise was combined with SSRIs ($n=268$, $\kappa=11$, $g = -0.55$, -0.86 to -0.23) or aerobic exercise was combined with psychotherapy ($n=404$, $\kappa=15$, $g = -0.54$, -0.76 to -0.32). All these treatments were significantly stronger than the standardised minimum clinically important difference compared with active control ($g = -0.20$), equating to an absolute g value of -1.16 . Dance, exercise combined with SSRIs, and walking or jogging

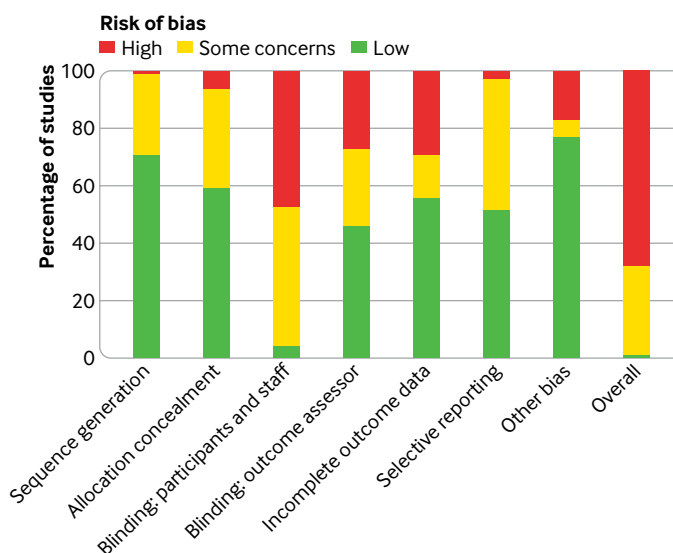


Fig 3 | Risk of bias summary plot showing percentage of included studies judged to be low, unclear, or high risk across Cochrane criteria for randomised trials

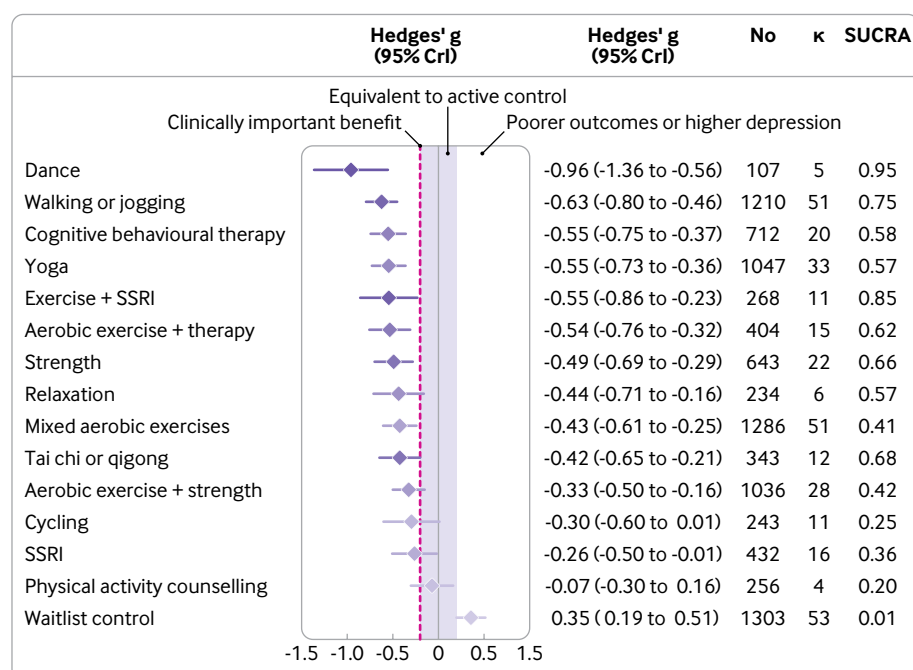


Fig 4 | Predicted effects of different exercise modalities on major depression compared with active controls (eg, usual care), with 95% credible intervals. The estimate of effects for the active control condition was a before and after change of Hedges' g of -0.95 (95% credible interval -1.10 to -0.79), n=3554, κ=113. Colour represents SUCRA from most likely to be helpful (dark purple) to least likely to be helpful (light purple). SSRI=selective serotonin reuptake inhibitor; SUCRA=surface under the cumulative ranking curve

were the treatments most likely to perform best when modelling the surface under the cumulative ranking curve (fig 4). For acceptability, the odds of participants dropping out of the study were lower for strength training (n=247, direct evidence κ=6, odds ratio 0.55, 95% credible interval 0.31 to 0.99) and yoga (n=264, κ=5, 0.57, 0.35 to 0.94) than for active control. The rate of dropouts was not significantly different from active control in any other arms (see supplementary file, section S8).

Consistent with other meta-analyses, effects were moderate for cognitive behaviour therapy alone (n=712, κ=20, g=-0.55, -0.75 to -0.37) and small for SSRIs (n=432, κ=16, g=-0.26, -0.50 to -0.01) compared with active controls (fig 4). These estimates are comparable to those of reviews that focused directly on psychotherapy (g=-0.67, -0.79 to -0.56)⁷ or pharmacotherapy (g=-0.30, -0.34 to -0.26).²⁵ However, our review was not designed to find all studies of these treatments, so these estimates should not usurp these directly focused systematic reviews.

Credibility assessment

Despite the large number of studies in the network, confidence in the effects were low (fig 5). This was largely due to the high within study bias described in the risk of bias summary plot. Reporting bias was also difficult to robustly assess because direct comparison with active control was often only provided in fewer than 10 studies. Many studies focused on one sex only, older adults, or those with comorbidities, so most arms

had some concerns about indirect comparisons. Credible intervals were seldom wide enough to change decision making, so concerns about imprecision were few. Heterogeneity did plausibly change some conclusions around clinical significance. Few studies showed problematic incoherence, meaning direct and indirect evidence usually agreed. Overall, walking or jogging had low confidence, with other modalities being very low.

Moderation by participant characteristics

The optimal modality appeared to be moderated by age and sex. Compared with models that only included exercise modality ($R^2=0.65$), R^2 was higher for models that included interactions with sex ($R^2=0.71$) and age ($R^2=0.69$). R^2 showed no substantial increase for models including baseline depression ($R^2=0.67$) or comorbidities ($R^2=0.66$; see supplementary file, section S9).

Effects appeared larger for women than men for strength training and cycling (fig 6). Effects appeared to be larger for men than women when prescribing yoga, tai chi, and aerobic exercise alongside psychotherapy. Yoga and aerobic exercise alongside psychotherapy appeared more effective for older participants than younger people (fig 7). Strength training appeared more effective when prescribed to younger participants than older participants. Some estimates were associated with substantial uncertainty because some modalities were not well studied in some groups (eg, tai chi for younger adults), and mean age of the sample was only available for 71% of the studies.

Active control versus	Direct comparisons (κ)	Within study bias	Reporting bias	Indirectness	Imprecision	Heterogeneity	Incoherence	Confidence rating
Aerobic exercise + strength	15	●	●	●	●	●	●	Very low
Aerobic exercise + therapy	6	●	●	●	●	●	●	Very low
Cycling	4	●	●	●	●	●	●	Very low
Dance	3	●	●	●	●	●	●	Very low
Mixed aerobic exercises	25	●	●	●	●	●	●	Very low
Physical activity counselling	3	●	●	●	●	●	●	Very low
Relaxation	1	●	●	●	●	●	●	Very low
Strength	10	●	●	●	●	●	●	Very low
Tai chi or qigong	7	●	●	●	●	●	●	Very low
Walking or jogging	23	●	●	●	●	●	●	Low
Yoga	13	●	●	●	●	●	●	Very low
Indirect evidence only								
Exercise + SSRI	0	●	●	●	●	●	●	Very low

Risk of bias
 ● Major concerns or high ● Some concerns or unclear ● No concerns or low

Fig 5 | Summary table for credibility assessment using confidence in network meta-analysis (CINeMA). SSRI=selective serotonin reuptake inhibitor

Moderation by intervention and design characteristics

Across modalities, a clear dose-response curve was observed for intensity of exercise prescribed (fig 8). Although light physical activity (eg, walking, hatha yoga) still provided clinically meaningful effects ($g=-0.58$, -0.82 to -0.33), expected effects were stronger for vigorous exercise (eg, running, interval training; $g=-0.74$, -1.10 to -0.38). This finding did not appear to be due to increased weekly energy expenditure: credible intervals were wide, which meant that the dose-response curve for METs/min prescribed per week was unclear (see supplementary file, section S10). Weak evidence suggested that shorter interventions (eg, 10 weeks: $g=-0.53$, -0.71 to -0.35) worked somewhat better than longer ones (eg, 30 weeks: $g=-0.37$, -0.79 to 0.03), with wide credible intervals again indicating high uncertainty (see supplementary file, section S11). We also moderated for the lag between the end of treatment and the measurement of the outcome. We found no indication that participants were likely to relapse within the measurement period (see supplementary file, section S12); effects remained steady when measured either directly after the intervention ($g=-0.59$, -0.80 to -0.39) or up to six months later ($g=-0.63$, -0.87 to -0.40).

Supplementary file section S13 provides coding for the behaviour change techniques and autonomy for each exercise arm. None of the behaviour change techniques significantly moderated overall effects. Contrary to expectations, studies describing a level of participant autonomy (ie, choice over frequency, intensity, type, or time) tended to show weaker effects ($g=-0.28$, -0.78 to 0.23) than those that did not ($g=-0.75$, -1.17 to -0.33 ; see supplementary file, section S14). This effect was consistent whether or not we included studies that used physical activity counselling (usually high autonomy).

Use of group exercise appeared to moderate the effects: although the overall effects were similar for individual ($g=-1.10$, -1.57 to -0.64) and group exercise ($g=-1.16$, -1.61 to -0.73), some interventions were better delivered in groups (yoga) and some were better delivered individually (strength training, mixed aerobic exercise; see supplementary file, section S15).

As preregistered, we tested whether study funding moderated effects. Models that included whether a study was funded did explain more variance ($R^2=0.70$) compared with models that included treatment alone ($R^2=0.65$). Funded studies showed stronger effects ($g=-1.01$, -1.19 to -0.82) than unfunded studies ($g=-0.77$, -1.09 to -0.46). We also moderated for the type of measure (self-report v clinician report). This did not explain a substantial amount of variance in the outcome ($R^2=0.66$).

Sensitivity analyses

Evidence of publication bias was found for overall estimates of exercise on depression compared with active controls, although not enough to nullify effects. The multilevel Egger's test showed significance ($F_{1,98}=23.93$, $P<0.001$). Funnel plots showed asymmetry, but the result of pooled effects remained statistically significant when only including non-significant studies (see supplementary file, section S16). No amount of publication bias would be sufficient to shrink effects to zero (s value=not possible). To reduce effects below clinical significance thresholds, studies with statistically significant results would need to be reported 58 times more frequently than studies with non-significant results.

Qualitative synthesis of mediation effects

Only a few of the studies used explicit mediation analyses to test hypothesised mechanisms of action.⁵⁴⁻⁵⁹

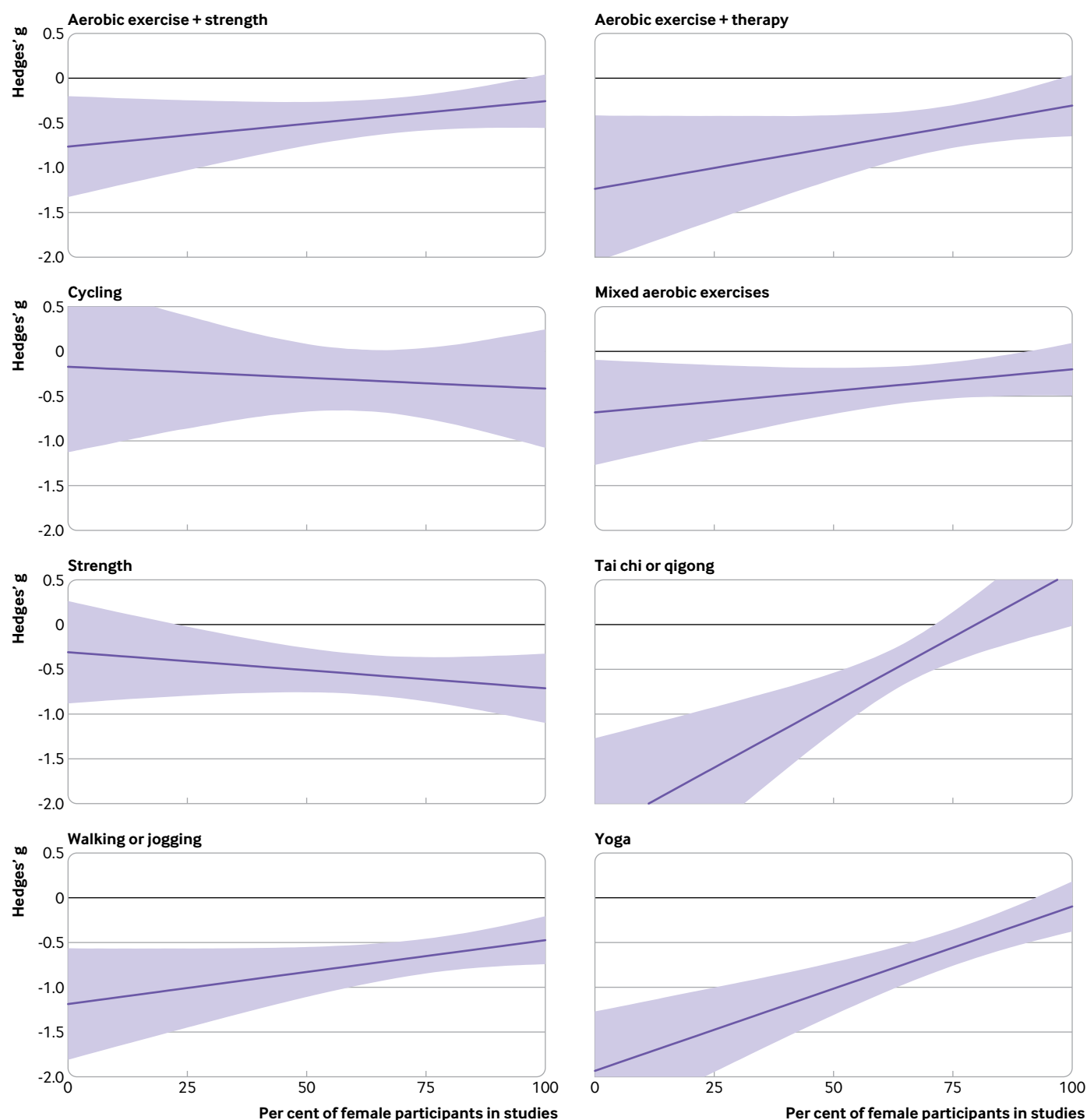


Fig 6 | Effects of interventions versus active control on depression (lower is better) by sex. Shading represents 95% credible intervals

One study found that both aerobic exercise and yoga led to decreased depression because participants ruminated less.⁵⁴ The study found that the effects of aerobic exercise (but not yoga) were mediated by increased acceptance.⁵⁴ “Perceived hassles” and awareness were not statistically significant mediators.⁵⁴ Another study found that the effects of yoga were mediated by increased self-compassion, but not rumination, self-criticism, tolerance of uncertainty, body awareness, body trust, mindfulness, and attentional biases.⁵⁵ One study found that the

effects from an aerobic exercise intervention were not mediated by long term physical activity, but instead were mediated by exercise specific affect regulation (eg, self-control for exercise).⁵⁷ Another study found that neither exercise self-efficacy nor depression coping self-efficacy mediated effects of aerobic exercise.⁵⁶ Effects of aerobic exercise were not mediated by the N2 amplitude from electroencephalography, hypothesised as a neuro-correlate of cognitive control deficits.⁵⁸ Increased physical activity did not appear to mediate the effects

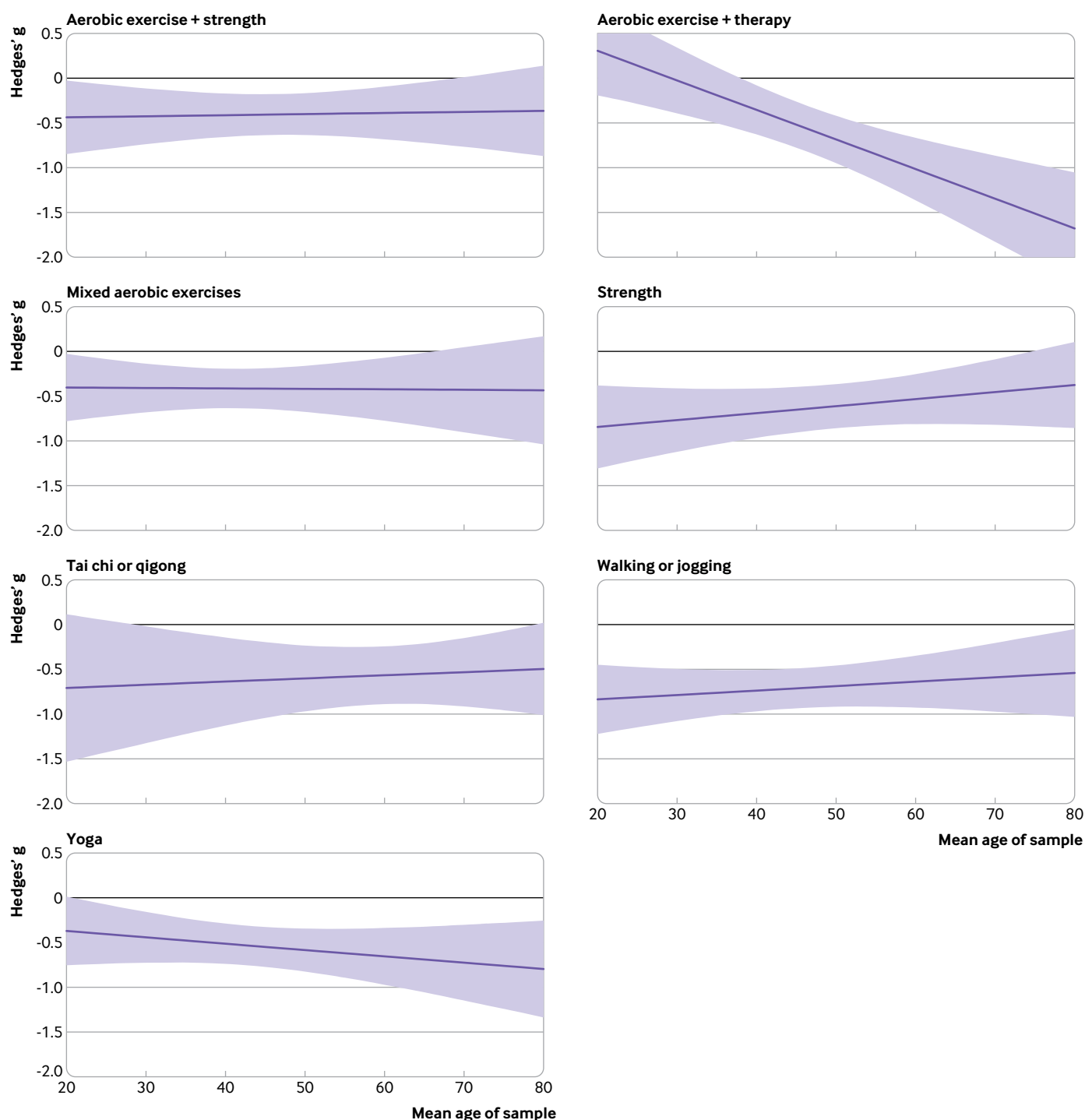


Fig 7 | Effects of interventions versus active control on depression (lower is better) by age. Shading represents 95% credible intervals

of physical activity counselling on depression.⁵⁹ It is difficult to infer strong conclusions about mechanisms on the basis of this small number of studies with low power.

Discussion

Summary of evidence

In this systematic review and meta-analysis of randomised controlled trials, exercise showed moderate effects on depression compared with

active controls, either alone or in combination with other established treatments such as cognitive behaviour therapy. In isolation, the most effective exercise modalities were walking or jogging, yoga, strength training, and dancing. Although walking or jogging were effective for both men and women, strength training was more effective for women, and yoga or qigong was more effective for men. Yoga was somewhat more effective among older adults, and strength training was more effective among

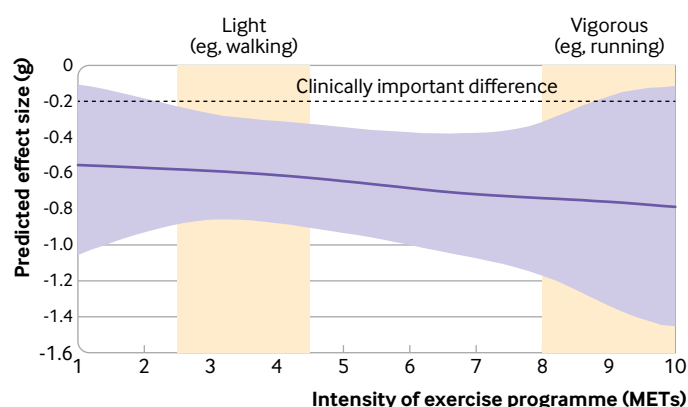


Fig 8 | Dose-response curve for intensity (METs) across exercise modalities compared with active control. METs=metabolic equivalents of task

younger people. The benefits from exercise tended to be proportional to the intensity prescribed, with vigorous activity being better. Benefits were equally effective for different weekly doses, for people with different comorbidities, or for different baseline levels of depression. Although confidence in many of the results was low, treatment guidelines may be overly conservative by conditionally recommending exercise as complementary or alternative treatment for patients in whom psychotherapy or pharmacotherapy is either ineffective or unacceptable.⁶⁰ Instead, guidelines for depression ought to include prescriptions for exercise and consider adapting the modality to participants' characteristics and recommending more vigorous intensity exercises.

Our review did not uncover clear causal mechanisms, but the trends in the data are useful for generating hypotheses. It is unlikely that any single causal mechanism explains all the findings in the review. Instead, we hypothesise that a combination of social interaction,⁶¹ mindfulness or experiential acceptance,⁶² increased self-efficacy,³³ immersion in green spaces,⁶³ neurobiological mechanisms,⁶⁴ and acute positive affect⁶⁵ combine to generate outcomes. Meta-analyses have found each of these factors to be associated with decreases in depressive symptoms, but no single treatment covers all mechanisms. Some may more directly promote mindfulness (eg, yoga), be more social (eg, group exercise), be conducted in green spaces (eg, walking), provide a more positive affect (eg, "runner's high"), or be more conducive to acute adaptations that may increase self-efficacy (eg, strength).⁶⁶ Exercise modalities such as running may satisfy many of the mechanisms, but they are unlikely to directly promote the mindful self-awareness provided by yoga and qigong. Both these forms of exercise are often practised in groups with explicit mindfulness but seldom have fast and objective feedback loops that improve self-efficacy. Adequately powered studies testing multiple mediators may help to focus more on understanding why exercise helps depression and less on whether exercise helps. We argue that understanding these mechanisms of action

is important for personalising prescriptions and better understanding effective treatments.

Our review included more studies than many existing reviews on exercise for depression.^{13 22 27 28} As a result, we were able to combine the strengths of various approaches to exercise and to make more nuanced and precise conclusions. For example, even taking conservative estimates (ie, the least favourable end of the credible interval), practitioners can expect patients to experience clinically significant effects from walking, running, yoga, qigong, strength training, and mixed aerobic exercise. Because we simultaneously assessed more than 200 studies, credible intervals were narrower than those in most existing meta-analyses.¹³ We were also able to explore non-linear relationships between outcomes and moderators, such as frequency, intensity, and time. These analyses supported some existing findings—for example, our study and the study by Heissel et al²² found that shorter interventions had stronger effects, at least for six months; our study and the study by Singh et al¹³ both found that effects were stronger with vigorous intensity exercise compared with light and moderate exercise. However, most existing reviews found various treatment modalities to be equally effective.^{13 27} In our review, some types of exercise had stronger effect sizes than others. We attribute this to the study level data available in a network meta-analysis compared with an overview of reviews²⁴ and higher power compared with meta-analyses with smaller numbers of included studies.^{22 28} Overviews of reviews have the ability to more easily cover a wider range of participants, interventions, and outcomes, but also risk double counting randomised trials that are included in separate meta-analyses. They often include heterogeneous studies without having as much control over moderation analyses (eg, Singh et al included studies covering both prevention and treatment¹³). Some of those reviews grouped interventions such as yoga with heterogeneous interventions such as stretching and qigong.¹³ This practise of combining different interventions makes it harder to interpret meta-analytical estimates. We used methods that enabled us to separately analyse the effects of these treatment modalities. In so doing, we found that these interventions do have different effects, with yoga being an intervention with strong effects and stretching being better described as an active control condition. Network meta-analyses revealed the same phenomenon with psychotherapy: researchers once concluded there was a dodo bird verdict, whereby "everybody has won, and all must have prizes,"⁶⁷ until network meta-analyses showed some interventions were robustly more effective than others.^{6 26}

Predictors of acceptability and outcomes

We found evidence to suggest good acceptability of yoga and strength training; although the measurement of study drop-out is an imperfect proxy of adherence. Participants may complete the study without doing

any exercise or may continue exercising and drop out of the study for other reasons. Nevertheless, these are useful data when considering adherence.

Behaviour change techniques, which are designed to increase adherence, did not meaningfully moderate the effect sizes from exercise. This may be due to several factors. It may be that the modality explains most of the variance between effects, such that behaviour change techniques (eg, presence or absence of feedback) did not provide a meaningful contribution. Many forms of exercise potentially contain therapeutic benefits beyond just energy expenditure. These characteristics of a modality may be more influential than coexisting behaviour change techniques. Alternatively, researchers may have used behaviour change techniques such as feedback or goal setting without explicitly reporting them in the study methods. Given the inherent challenges of behaviour change among people with depression,²⁹ and the difficulty in forecasting which strategies are likely to be effective,⁶⁸ we see the identification of effective techniques as important.

We did find that autonomy, as provided in the methods of included studies, predicted effects, but in the opposite direction to our hypotheses: more autonomy was associated with weaker effects. Physical activity counselling, which usually provides a great deal of patient autonomy, was among the lowest effect sizes in our meta-analysis. Higher autonomy judgements were associated with weaker outcomes regardless of whether physical activity counselling was included in the model. One explanation for these data is that people with depression benefit from the clear direction and accountability of a standardised prescription. When provided with more freedom, the low self-efficacy that is symptomatic of depression may stop patients from setting an appropriate level of challenge (eg, they may be less likely to choose vigorous exercise). Alternatively, participants were likely autonomous when self-selecting into trials with exercise modalities they enjoyed, or those that fit their social circumstances. After choosing something value aligned, autonomy within the trial may not have helped. Either way, data should be interpreted with caution. Our judgement of the autonomy provided in the methods may not reflect how much autonomy support patients actually felt. The patient's perceived autonomy is likely determined by a range of factors not described in the methods (eg, the social environment created by those delivering the programme, or their social identity), so other studies that rely on patient reports of the motivational climate are likely to be more reliable.³³ Our findings reiterate the importance of considering these patient reports in future research of exercise for depression.

Our findings suggest that practitioners could advocate for most patients to engage in exercise. Those patients may benefit from guidance on intensity (ie, vigorous) and types of exercise that appear to work well (eg, walking, running, mixed aerobic exercise, strength training, yoga, tai chi, qigong) and be well

tolerated (eg, strength training and yoga). If social determinants permit,⁶⁶ engaging in group exercise or structured programmes could provide support and guidance to achieve better outcomes. Health services may consider offering these programmes as an alternative or adjuvant treatment for major depression. Specifically, although the confidence in the evidence for exercise is less strong than for cognitive behavioural therapy, the effect sizes seem comparable, so it may be an alternative for patients who prefer not to engage in psychotherapy. Previous reviews on those with mild-moderate depression have found similar effects for exercise or SSRIs, or the two combined.^{13 14} In contrast, we found some forms of exercise to have stronger effects than SSRIs alone. Our findings are likely related to the larger power in our review ($n=14\,170$) compared with previous reviews (eg, $n=2551$),¹⁴ and our ability to better account for heterogeneity in exercise prescriptions. Exercise may therefore be considered a viable alternative to drug treatment. We also found evidence that exercise increases the effects of SSRIs, so offering exercise may act as an adjuvant for those already taking drugs. We agree with consensus statements that professionals should still account for patients' values, preferences, and constraints, ensuring there is shared decision making around what best suits the patient.⁶⁶ Our review provides data to help inform that decision.

Strengths, limitations, and future directions

Based on our findings, dance appears to be a promising treatment for depression, with large effects found compared with other interventions in our review. But the small number of studies, low number of participants, and biases in the study designs prohibits us from recommending dance more strongly. Given most research for the intervention has been in young women (88% female participants, mean age 31 years), it is also important for future research to assess the generalisability of the effects to different populations, using robust experimental designs.

The studies we found may be subject to a range of experimental biases. In particular, researchers seldom blinded participants or staff delivering the intervention to the study's hypotheses. Blinding for exercise interventions may be harder than for drugs²³; however, future studies could attempt to blind participants and staff to the study's hypotheses to avoid expectancy effects.⁶⁹ Some of our ratings are for studies published before the proliferation of reporting checklists, so the ratings might be too critical.²³ For example, before CONSORT, few authors explicitly described how they generated a random sequence.²³ Therefore, our risk of bias judgements may be too conservative. Similarly, we planned to use the Cochrane risk of bias (RoB) 1 tool⁴⁰ so we could use the most recent Cochrane review of exercise and depression¹² to calibrate our raters, and because RoB 2 had not yet been published.⁷⁰ Although assessments of bias between the two tools are generally comparable,⁷¹ the RoB 1 tool can be more conservative when assessing open label studies with subjective

assessments (eg, unblinded studies with self-reported measures for depression).⁷¹ As a result, future reviews should consider using the latest risk of bias tool, which may lead to different assessments of bias in included studies.

Most of the main findings in this review appear robust to risks from publication bias. Specifically, pooled effect sizes decreased when accounting for risk of publication bias, but no degree of publication bias could nullify effects. We did not exclude grey literature, but our search strategy was not designed to systematically search grey literature or trial registries. Doing so can detect additional eligible studies⁷² and reveal the numbers of completed studies that remain unpublished.⁷³ Future reviews should consider more systematic searches for this kind of literature to better quantify and mitigate risk of publication bias.

Similarly, our review was able to integrate evidence that directly compared exercise with other treatment modalities such as SSRIs or psychotherapy, while also informing estimates using indirect evidence (eg, comparing the relative effects of strength training and SSRIs when tested against a waitlist control). Our review did not, however, include all possible sources of indirect evidence. Network meta-analyses exist that directly focus on psychotherapy⁷ and pharmacotherapy,²⁵ and these combined for treating depression.⁶ Those reviews include more than 500 studies comparing psychological or drug interventions with controls. Harmonising the findings of those reviews with ours would provide stronger data on indirect effects.

Our review found some interesting moderators by age and sex, but these were at the study level rather than individual level—that is, rather than being able to determine whether women engaging in a strength intervention benefit more than men, we could only conclude that studies with more women showed larger effects than studies with fewer women. These studies may have been tailored towards women, so effects may be subject to confounding, as both sex and intervention may have changed. The same finding applied to age, where studies on older adults were likely adapted specifically to this age group. These between study differences may explain the heterogeneity in the effects of interventions, and confounding means our moderators for age and sex should be interpreted cautiously. Future reviews should consider individual patient meta-analyses to allow for more detailed assessments of participant level moderators.

Finally, for many modalities, the evidence is derived from small trials (eg, the median number of walking or jogging arms was 17). In addition to reducing risks from bias, primary research may benefit from deconstruction designs or from larger, head-to-head analyses of exercise modalities to better identify what works best for each candidate.

Clinical and policy implications

Our findings support the inclusion of exercise as part of clinical practice guidelines for depression, particularly

vigorous intensity exercise. Doing so may help bridge the gap in treatment coverage by increasing the range of first line options for patients and health systems.⁹ Globally there has been an attempt to reduce stigma associated with seeking treatment for depression.⁷⁴ Exercise may support this effort by providing patients with treatment options that carry less stigma. In low resource or funding constrained settings, group exercise interventions may provide relatively low cost alternatives for patients with depression and for health systems. When possible, ideal treatment may involve individualised care with a multidisciplinary team, where exercise professionals could take responsibility for ensuring the prescription is safe, personalised, challenging, and supported. In addition, those delivering psychotherapy may want to direct some time towards tackling cognitive and behavioural barriers to exercise. Exercise professionals might need to be trained in the management of depression (eg, managing risk) and to be mindful of the scope of their practice while providing support to deal with this major cause of disability.

Conclusions

Depression imposes a considerable global burden. Many exercise modalities appear to be effective treatments, particularly walking or jogging, strength training, and yoga, but confidence in many of the findings was low. We found preliminary data that may help practitioners tailor interventions to individuals (eg, yoga for older men, strength training for younger women). The World Health Organization recommends physical activity for everyone, including those with chronic conditions and disabilities,⁷⁵ but not everyone can access treatment easily. Many patients may have physical, psychological, or social barriers to participation. Still, some interventions with few costs, side effects, or pragmatic barriers, such as walking and jogging, are effective across people with different personal characteristics, severity of depression, and comorbidities. Those who are able may want to choose more intense exercise in a structured environment to further decrease depression symptoms. Health systems may want to provide these treatments as alternatives or adjuvants to other established interventions (cognitive behaviour therapy, SSRIs), while also attenuating risks to physical health associated with depression.³ Therefore, effective exercise modalities could be considered alongside those intervention as core treatments for depression.

AUTHOR AFFILIATIONS

¹School of Psychology, University of Queensland, St Lucia, QLD 4072, Australia

²Institute for Positive Psychology and Education, Australian Catholic University, North Sydney, NSW, Australia

³Department of Physical Education and Sport, University of Seville, Seville, Spain

⁴School of Health and Behavioural Sciences, Australian Catholic University, Strathfield, NSW, Australia

⁵Department of Clinical Biomechanics and Sports Science, University of Southern Denmark, Odense, Denmark

⁶Biomedical Research and Innovation Institute of Cádiz (INIBICA) Research Unit, University of Cádiz, Spain

⁷School of Health and Behavioural Sciences, University of the Sunshine Coast, Petrie, QLD, Australia

⁸School of Education, University of Newcastle, Callaghan, NSW, Australia

⁹School of Health and Behavioural Sciences, Australian Catholic University, Banyo, QLD, Australia

¹⁰School of Education, Australian Catholic University, Strathfield, NSW, Australia

¹¹Australian Institute of Health Innovation, Macquarie University, Macquarie Park, NSW, Australia

¹²Children's Hospital Westmead Clinical School, University of Sydney, Westmead, NSW, Australia

¹³Australian Catholic University, North Sydney, NSW, Australia

¹⁴Centre for Health Research, University of Southern Queensland, Springfield, QLD, Australia

¹⁵Faculty of Sport and Health Science, University of Jyväskylä, Jyväskylä, Finland

We thank Lachlan McKee for his assistance with data extraction. We also thank Juliette Grosvenor and another librarian (anonymous) for their review of our search strategy.

Contributors: MN led the project, drafted the manuscript, and is the guarantor. MN, TS, PT, MM, BdPC, PP, SB, and CL drafted the initial study protocol. MN, TS, PT, BdPC, DvdH, JS, MM, RP, LP, RV, HA, and BV conducted screening, extraction, and risk of bias assessment. MN, JS, and JM coded methods for behaviour change techniques. MN and DGG conducted statistical analyses. PP, SB, and CL provided supervision and mentorship. All authors reviewed and approved the final manuscript. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: None received.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: Data and code for reproducing analyses are available on the Open Science Framework (<https://osf.io/nzw6u/>).

The lead author (MN) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Dissemination to participants and related patient and public communities: We plan to disseminate the findings of this study to lay audiences through mainstream and social media.

Provenance and peer review: Not commissioned; externally peer reviewed.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

- World Health Organization. Depression. 2020 [cited 2020 Mar 12]. <https://www.who.int/news-room/fact-sheets/detail/depression>
- Birkjær M, Kaats M, Rubio A. Wellbeing adjusted life years: A universal metric to quantify the happiness return on investment. Happiness Research Institute; 2020. <https://www.happinessresearchinstitute.com/waly-report>
- Rugulies R. Depression as a predictor for coronary heart disease. a review and meta-analysis. *Am J Prev Med* 2002;23:51-61. doi:10.1016/S0749-3797(02)00439-7
- Jacobson NC, Newman MG. Anxiety and depression as bidirectional risk factors for one another: A meta-analysis of longitudinal studies. *Psychol Bull* 2017;143:1155-200. doi:10.1037/bul0000111
- Pinquart M, Duberstein PR. Depression and cancer mortality: a meta-analysis. *Psychol Med* 2010;40:1797-810. doi:10.1017/S0033291709992285
- Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry* 2020;19:92-107. doi:10.1002/wps.20701
- Cuijpers P, Quero S, Noma H, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry* 2021;20:283-93. <https://onlinelibrary.wiley.com/doi/10.1002/wps.20860>
- Strawbridge R, Carter B, Marwood L, et al. Augmentation therapies for treatment-resistant depression: systematic review and meta-analysis. *Br J Psychiatry* 2019;214:42-51. doi:10.1192/bjp.2018.233
- Moitra M, Santomauro D, Collins PY, et al. The global gap in treatment coverage for major depressive disorder in 84 countries from 2000-2019: A systematic review and Bayesian meta-regression analysis. *PLoS Med* 2022;19:e1003901. doi:10.1371/journal.pmed.1003901
- Hu MX, Turner D, General E, et al. Exercise interventions for the prevention of depression: a systematic review of meta-analyses. *BMC Public Health* 2020;20:1255. doi:10.1186/s12889-020-09323-y
- Rimer J, Dwan K, Lawlor DA, et al. Exercise for depression. *Cochrane Database Syst Rev* 2012;(7):CD004366. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84866461883&partnerID=40&md5=017f47a207b41fc4d2fddb12b69e6ba2>
- Cooney GM, Dwan K, Greig CA, et al. Exercise for depression. *Cochrane Database Syst Rev* 2013;2013:CD004366. doi:10.1002/14651858.CD004366.pub6
- Singh B, Olds T, Curtis R, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. *Br J Sports Med* 2023;57:1203-9. doi:10.1136/bjsports-2022-106195
- Recchia F, Leung CK, Chin EC, et al. Comparative effectiveness of exercise, antidepressants and their combination in treating non-severe depression: a systematic review and network meta-analysis of randomised controlled trials. *Br J Sports Med* 2022;56:1375-80. doi:10.1136/bjsports-2022-105964
- Owen PJ, Miller CT, Mundell NL, et al. Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis. *Br J Sports Med* 2020;54:1279-87. doi:10.1136/bjsports-2019-100886
- Gallardo-Gómez D, Del Pozo-Cruz J, Noetel M, Álvarez-Barbosa F, Alfonso-Rosa RM, Del Pozo Cruz B. Optimal dose and type of exercise to improve cognitive function in older adults: A systematic review and bayesian model-based network meta-analysis of RCTs. *Ageing Res Rev* 2022;76:101591. doi:10.1016/j.arr.2022.101591
- Naci H, Salcher-Konrad M, Dias S, et al. How does exercise treatment compare with antihypertensive medications? A network meta-analysis of 391 randomised controlled trials assessing exercise and medication effects on systolic blood pressure. *Br J Sports Med* 2019;53:859-69. doi:10.1136/bjsports-2018-099921
- National Collaborating Centre for Mental Health (UK). Depression: The Treatment and Management of Depression in Adults (Updated Edition). Leicester (UK): British Psychological Society; <https://www.ncbi.nlm.nih.gov/pubmed/22132433>
- Malhi GS, Bell E, Bassett D, et al. The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry* 2021;55:7-117. doi:10.1177/0004867420979353
- American Psychiatric Association. Practice Guideline for the Treatment of Patients with Major Depressive Disorder. Third Edition. Washington, DC: American Psychiatric Association; 2010. 87 p. https://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd-1410197717630.pdf
- NICE. Depression in adults: treatment and management. [cited 2023 Mar 13]. National Institute for Health and Care Excellence; 2022 <https://www.nice.org.uk/guidance/ng222/resources>
- Heissel A, Heinen D, Brokmeier LL, et al. Exercise as medicine for depressive symptoms? A systematic review and meta-analysis with meta-regression. *Br J Sports Med* 2023;57:1049-57. doi:10.1136/bjsports-2022-106282
- Ekkekakis P. Honey, I shrunk the pooled SMD! Guide to critical appraisal of systematic reviews and meta-analyses using the Cochrane review on exercise for depression as example. *Ment Health Phys Act* 2015;8:21-36. <https://www.sciencedirect.com/science/article/pii/S1755296614000593>doi:10.1016/j.mhpa.2014.12.001
- Chaimani A, Caldwell DM, Li T, Higgins JPT, Salanti G. Undertaking network meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane; 2022. www.training.cochrane.org/handbook
- Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet* 2018;391:1357-66. doi:10.1016/S0140-6736(17)32802-7
- Barth J, Munder T, Gerger H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med* 2013;10:e1001454. doi:10.1371/journal.pmed.1001454

- 27 Miller KJ, Gonçalves-Bradley DC, Areerob P, Hennessy D, Mesagno C, Grace F. Comparative effectiveness of three exercise types to treat clinical depression in older adults: A systematic review and network meta-analysis of randomised controlled trials. *Ageing Res Rev* 2020;58:100999. doi:10.1016/j.arr.2019.100999
- 28 Yu Q, Wong KK, Lei OK, et al. Comparative Effectiveness of Multiple Exercise Interventions in the Treatment of Mental Health Disorders: A Systematic Review and Network Meta-Analysis. *Sports Med Open* 2022;8:135. doi:10.1186/s40798-022-00529-5
- 29 Glowacki K, Duncan MJ, Gainforth H, Faulkner G. Barriers and facilitators to physical activity and exercise among adults with depression: A scoping review. *Ment Health Phys Act* 2017;13:108-19. <https://www.sciencedirect.com/science/article/pii/S1755296617300959>doi:10.1016/j.mhpa.2017.10.001.
- 30 Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med* 2013;46:81-95. doi:10.1007/s12160-013-9486-6
- 31 Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. *Health Psychol* 2009;28:690-701. doi:10.1037/a0016136
- 32 French DP, Olander EK, Chisholm A, Mc Sharry J. Which behaviour change techniques are most effective at increasing older adults' self-efficacy and physical activity behaviour? A systematic review. *Ann Behav Med* 2014;48:225-34. <https://www.ncbi.nlm.nih.gov/pubmed/24648017>. doi:10.1007/s12160-014-9593-z
- 33 Ntoumanis N, Ng JYY, Prestwich A, et al. A meta-analysis of self-determination theory-informed intervention studies in the health domain: effects on motivation, health behavior, physical, and psychological health. *Health Psychol Rev* 2021;15:214-44. doi:10.1080/17437199.2020.1718529
- 34 Hutton B, Salanti G, Caldwell DM, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med* 2015;162:777-84. <https://www.acpjournals.org/doi/abs/10.7326/M14-2385>. doi:10.7326/M14-2385
- 35 Nikolakopoulou A, Higgins JPT, Papakonstantinou T, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med* 2020;17:e1003082. doi:10.1371/journal.pmed.1003082
- 36 Beck AT, Steer RA, Brown GK. *BDI-II, Beck Depression Inventory: Manual*. Psychological Corporation, 1996.
- 37 Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep* 1985;100:126-31. <https://www.ncbi.nlm.nih.gov/pubmed/3920711>.
- 38 Veritas Health Innovation. Covidence systematic review software. Melbourne, Australia; 2023. www.covidence.org
- 39 Ainsworth BE, Haskell WL, Herrmann SD, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. *Med Sci Sports Exerc* 2011;43:1575-81. doi:10.1249/MSS.0b013e31821e12e2
- 40 Higgins JPT, Altman DG, Gøtzsche PC, et al. Cochrane Methods Group. Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928. doi:10.1136/bmj.d5928
- 41 Lin L, Zhang J, Hodges JS, Chu H. Performing Arm-Based Network Meta-Analysis in R with the pnetmeta Package. *J Stat Softw* 2017;80:5. doi:10.18637/jss.v080.i05
- 42 Dias S, Welton NJ, Sutton AJ, Ades AE. NICE DSU technical support document 2: a generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. In: National Institute for Health and Care Excellence (NICE), editor. NICE Decision Support Unit Technical Support Documents. London: Citeseer; 2011. <https://www.ncbi.nlm.nih.gov/books/NBK310366/>
- 43 Faltinsen E, Todorovac A, Staxen Bruun L, et al. Control interventions in randomised trials among people with mental health disorders. *Cochrane Database Syst Rev* 2022;4:MR000050. doi:10.1002/14651858.MR000050.pub2
- 44 R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2022. <https://www.R-project.org/>
- 45 Bürkner PC. brms: An R Package for Bayesian Multilevel Models Using Stan. *J Stat Softw* 2017;80:1-28. <https://www.jstatsoft.org/article/view/v080i01>doi:10.18637/jss.v080.i01.
- 46 Harrer M, Cuijpers P, Furukawa TA, Ebert DD. *Doing meta-analysis with R: A hands-on guide*. Chapman and Hall, 2021, doi:10.1201/9781003107347.
- 47 Hengartner MP, Plödel M. Estimates of the minimal important difference to evaluate the clinical significance of antidepressants in the acute treatment of moderate-to-severe depression. *BMJ Evid Based Med* 2022;27:69-73. doi:10.1136/bmjebm-2020-116000
- 48 Balduzzi S, Rücker G, Nikolakopoulou A, et al. netmeta: An R Package for Network Meta-Analysis Using Frequentist Methods. *J Stat Softw* 2023;106:1-40. <https://www.jstatsoft.org/index.php/jss/article/view/4487>doi:10.18637/jss.v106.i02.
- 49 Papakonstantinou T, Nikolakopoulou A, Higgins JPT, Egger M, Salanti G. CINeMA: Software for semiautomated assessment of the confidence in the results of network meta-analysis. *Campbell Syst Rev* 2020;16:e1080. doi:10.1002/cl2.1080
- 50 Li X, Dusseldorp E, Su X, Meulman JJ. Multiple moderator meta-analysis using the R-package Meta-CART. *Behav Res Methods* 2020;52:2657-73. doi:10.3758/s13428-020-01360-0
- 51 Sterne JAC, Sutton AJ, Ioannidis JPA, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ* 2011;343:d4002. doi:10.1136/bmj.d4002
- 52 Mathur MB, VanderWeele TJ. Sensitivity analysis for publication bias in meta-analyses. *J R Stat Soc Ser C Appl Stat* 2020;69:1091-119. doi:10.1111/rssc.12440
- 53 Leung LYL, Tam HL, Ho JKM. Effectiveness of Tai Chi on older adults: A systematic review of systematic reviews with re-meta-analysis. *Arch Gerontol Geriatr* 2022;103:104796. doi:10.1016/j.archger.2022.104796
- 54 La Rocque CL, Mazurka R, Stuckless TJR, Pyke K, Harkness KL. Randomized controlled trial of bikram yoga and aerobic exercise for depression in women: Efficacy and stress-based mechanisms. *J Affect Disord* 2021;280(Pt A):457-66. doi:10.1016/j.jad.2020.10.067
- 55 Vollbrecht NK, Hoenders HJR, Bartels-Velthuis AA, et al. Mindful yoga intervention as add-on to treatment as usual for young women with major depressive disorder: Results from a randomized controlled trial. *J Consult Clin Psychol* 2022;90:925-41. <https://libkey.io/libraries/590/openurl?genre=article&aulast=Vollbrecht&issn=0022-006X&title=Journal+of+Consulting+and+C>. doi:10.1037/ccp0000777
- 56 Chu IH. Effect of exercise intensity during aerobic training on depressive symptoms in initially sedentary depressed women. *Dissertation Abstracts International B: Sciences and Engineering* 2018;78(11-B(E)).
- 57 Zeigbig JM, Seiffer B, Frei AK, et al. Long-term efficacy of exercise across diagnostically heterogeneous mental disorders and the mediating role of affect regulation skills. *Psychol Sport Exerc* 2023;64:102340. doi:10.1016/j.psychsport.2022.102340
- 58 Olson RL, Brush CJ, Ehmann PJ, Alderman BL. A randomized trial of aerobic exercise on cognitive control in major depression. *Clin Neurophysiol* 2017;128:903-13. doi:10.1016/j.clinph.2017.01.023
- 59 Bombardier CH, Ehde DM, Gibbons LE, et al. Telephone-based physical activity counseling for major depression in people with multiple sclerosis. *J Consult Clin Psychol* 2013;81:89-99. doi:10.1037/a0031242
- 60 American Psychological Association. Clinical practice guideline for the treatment of depression across three age cohorts. American Psychological Association; 2019. <https://www.apa.org/depression-guideline/>
- 61 Park M, Cuijpers P, van Straten A, Reynolds CF3rd. The effects of psychotherapy for adult depression on social support: A meta-analysis. *Cognit Ther Res* 2014;38:600-11. doi:10.1007/s10608-014-9630-z
- 62 Johannsen M, Nissen ER, Lundorff M, O'Toole MS. Mediators of acceptance and mindfulness-based therapies for anxiety and depression: A systematic review and meta-analysis. *Clin Psychol Rev* 2022;94:102156. doi:10.1016/j.cpr.2022.102156
- 63 Coventry PA, Brown JE, Pervin J, et al. Nature-based outdoor activities for mental and physical health: Systematic review and meta-analysis. *SSM Popul Health* 2021;16:100934. <https://www.sciencedirect.com/science/article/pii/S2352827321002093>. doi:10.1016/j.ssmph.2021.100934
- 64 Schuch FB, Deslandes AC, Stubbs B, Gosmann NP, Silva CT, Fleck MP de A. Neurobiological effects of exercise on major depressive disorder: A systematic review. *Neurosci Biobehav Rev* 2016;61:1-11. doi:10.1016/j.neubiorev.2015.11.012
- 65 Niven A, Laird Y, Saunders DH, Phillips SM. A systematic review and meta-analysis of affective responses to acute high intensity interval exercise compared with continuous moderate- and high-intensity exercise. *Health Psychol Rev* 2021;15:540-73. doi:10.1080/17437199.2020.1728564
- 66 Vella SA, Aidman E, Teychenne M, et al. Optimising the effects of physical activity on mental health and wellbeing: A joint consensus statement from Sports Medicine Australia and the Australian Psychological Society. *J Sci Med Sport* 2023;26:132-9. doi:10.1016/j.jsams.2023.01.001
- 67 Hunsley J, Di Giulio G. Dodo bird, phoenix, or urban legend? *Sci Rev Ment Health Pract* 2002;1:11-22. <http://www.srmhp.org/0101/psychotherapy-equivalence.html>.
- 68 Milkman KL, Gromet D, Ho H, et al. Megastudies improve the impact of applied behavioural science. *Nature* 2021;600:478-83. <https://www.nature.com/articles/s41586-021-04128-4>. doi:10.1038/s41586-021-04128-4

- 69 Hecksteden A, Faude O, Meyer T, Donath L. How to Construct, Conduct and Analyze an Exercise Training Study? *Front Physiol* 2018;9:1007. doi:10.3389/fphys.2018.01007
- 70 Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:l4898. doi:10.1136/bmj.l4898
- 71 Richter B, Hemmingsen B. Comparison of the Cochrane risk of bias tool 1 (RoB 1) with the updated Cochrane risk of bias tool 2 (RoB 2). Cochrane; 2021. Report No.: 1. https://community.cochrane.org/sites/default/files/uploads/inline-files/RoB1_2_project_220529_BR%20KK%20formatted.pdf
- 72 Lefebvre C, Glanville J, Briscoe S, et al. Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, et al, eds. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 2022, www.training.cochrane.org/handbook.
- 73 Coronado-Montoya S, Levis AW, Kwakkenbos L, Steele RJ, Turner EH, Thoms BD. Reporting of positive results in randomized controlled trials of mindfulness-based mental health interventions. *PLoS One* 2016;11:e0153220. doi:10.1371/journal.pone.0153220
- 74 Angermeyer MC, van der Auwera S, Carta MG, Schomerus G. Public attitudes towards psychiatry and psychiatric treatment at the beginning of the 21st century: a systematic review and meta-analysis of population surveys. *World Psychiatry* 2017;16:50-61. doi:10.1002/wps.20383
- 75 Bull FC, Al-Ansari SS, Biddle S, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. *Br J Sports Med* 2020;54:1451-62. doi:10.1136/bjsports-2020-102955

Supplementary information: Supplementary files S0-S16