ROUND THE CORNER

Exercise for depression: strong option with clinical benefits†

Lena Lisa Maria Beckers, Daniela Almeida Borges & Brennan Delattre

SUMMARY

Exercise, investigated as a treatment option for depression, is promising in that it has previously shown antidepressant effects and is accessible, low-cost and additionally provides physical health benefits. This commentary assesses the effects of exercise interventions relative to control, psychological and pharmacological treatments on depressive symptoms, as reported in Clegg et al's (2026) recent Cochrane Review. The review suggests that exercise may reduce depressive symptoms and provide benefits similar to those of psychological and pharmacological treatments. This must be interpreted with caution in light of uncertain evidence, but results show that further investigation is warranted.

KEYWORDS

Depressive disorders; physical exercise intervention; sports and exercise psychiatry; antidepressants; psychological treatments.

Major depressive disorder (MDD) is characterised by persistent low mood and reduced interest or pleasure in activities, often accompanied by changes in sleep, appetite, cognition and fatigue (American Psychiatric Association 2013). It has been associated with increased morbidity and mortality, multiple physical comorbidities and substantial impairments in physical, mental and social functioning (Marx 2023). About 5.2% of adults worldwide are estimated to suffer from depression, and this number is rising (World Health Organization 2026).

Although effective treatments exist, important barriers to treatment uptake and efficacy remain. Pharmacological treatments, such as selective serotonin reuptake inhibitors (SSRIs), often require several weeks before providing clinically meaningful benefits and can have adverse side-effects (Rush 2009) that, along with social stigma and a preference for non-pharmacological treatments, can hamper adherence (Meng 2024). Furthermore, only about one-third of patients respond to first-line antidepressant treatments, and clinicians cannot yet reliably predict who will benefit from which intervention (Cuijpers 2020). Psychological treatments such as cognitive–behavioural therapy (CBT) are also limited by accessibility, motivation and willingness to engage with treatment, which may be negatively affected by social stigma or the risk of increased emotional distress (Wittmann 2023). These limitations highlight the need for alternative treatment options.

Exercise interventions

Lifestyle interventions targeting diet, exercise or sleep have gained more attention as potential alternative treatment options for depression (Gonzalez-Garcia 2026). Exercise interventions have been associated with antidepressant effects (e.g. Ross 2023). The mechanisms are not yet clear, but biopsychosocial pathways have been proposed. Biologically, exercise modulates neurotransmitter and brain-derived neurotrophic factor (BDNF) levels (Ross 2023) and reduces inflammation, the last of which is associated with increased dopaminergic signalling and thereby thought to improve anhedonia, fatigue and cognitive impairment (Hird 2024). Exercise also acutely reduces rumination (Welkerling 2026), highlighting a possible psychological pathway. Finally, certain exercise types might improve mood through fostering social connectedness, a protective factor against depression (e.g. Delattre 2024) and thus a potential social pathway.

Exercise is more accessible than many other treatment options (Box 1), generally low-cost and additionally provides physical health benefits that counter common physical comorbidities associated with depression (such as cardiovascular and metabolic disease; Ross 2023). Hence, it may be a useful option for patients who do not respond to, cannot access or prefer not to receive conventional treatments, and it has been argued that not prescribing exercise may even constitute malpractice (Fabiano 2025).

The Cochrane Review

In their Cochrane Review, Clegg et al (2026) evaluated the effectiveness of exercise for treating depression compared with control conditions and

Lena Lisa Maria Beckers, MSc, is a research assistant in the Department of Psychiatry, University of Oxford, and she is affiliated to the National Institute for Health and Care Research Biomedical Research Centre: Oxford Health, Oxford, UK. She conducts research into the neuropsychological mechanisms of depression and antidepressant action, using a combination of ultra-high field neuroimaging and psychopharmacology.

Daniela Almeida Borges, MRCPsych, is a Year 6 Specialty Trainee (ST6) in general and liaison psychiatry with Oxford University Hospitals NHS Foundation Trust, and an MSc by Research student in the Department of Psychiatry, University of Oxford, UK, researching the neuropsychological mechanisms of orexin antagonists. She holds honorary contracts for research with Oxford Health NHS Foundation Trust, Oxford, UK, and the Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

Brennan Delattre, MSc, is a doctoral candidate in the Department of Psychiatry, University of Oxford, Oxford, UK. Her research sits at the intersection of social psychology and intervention science, and she examines the efficacy of social dance and creative, movement-based activities for individuals suffering from depression and related mental disorders within and outside of public healthcare settings.

Correspondence Lena Lisa Maria Beckers. Email: lena.beckers@psych.ox.ac.uk

First received 9 Jul 2026 UTC
Revised 16 Aug 2026 UTC
Accepted 7 Sep 2026 UTC

Copyright and usage
© The Author(s), 2026. Published by Cambridge University Press on behalf of Royal College of Psychiatrists.

†Commentary on . . . Exercise for depression (Cochrane Corner). See this issue.

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

BOX 1 Accessibility of interventions

A major consideration when implementing interventions for depression is their accessibility, both for the healthcare providers prescribing them and the patients undertaking them.

Health equity

Health equity is defined as the absence of avoidable and unfair differences in health (Welch 2024), which may be experienced across any of the following components, collectively referred to as PROGRESS-Plus: place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, social capital and other characteristics (e.g. sexual orientation, age, disability) (O'Neill 2013). For this to be meaningfully considered in research when assessing different interventions, we need trials that consider equity issues. For instance, this includes the collection of health equity metrics; seeking to recruit diverse population groups; determining the cost of treatment and intervention availability for providers; and/or directing patients towards resources. This is imperative for establishing overall accessibility and equitable implementation of an intervention at a population level but is often not reported in trial results.

Despite Cochrane's emphasis on the importance of including health equity metrics, this review (Clegg 2026) was unable to directly address equity issues owing to a combination of precedent set by previous versions of the review and health inequalities not being addressed in the source studies. The review authors report that most trials included did not report ensuring a representation of different groups in the population beyond a depressed sample. Moreover, the review did not assess costs of exercise treatment, as this was not reported in any of the included trials. Echoing Clegg et al's suggestion, future trials considering equity issues and treatment cost are needed.

Beyond this, specific accessibility considerations relevant to the nature of the intervention are detailed below. Please note that this list is not exhaustive and seeks to provide a starting point when considering the accessibility of an intervention, with examples associated with the context of exercise for depression.

Participant considerations

General access: Barriers to participation must be considered. Access to treatment for depression varies widely between socioeconomic classes and global regions for a host of reasons (for reviews, see Qin 2020; Henking 2023; Marx 2023). To use an example, does being able to participate in one's chosen intervention involve filling out forms, meeting with a healthcare provider virtually or in person, picking up prescriptions at a pharmacy, regularly scheduled sessions or having access to specialised gear or equipment? The considerations are different for pharmacological approaches, talking therapies and lifestyle interventions, but the burden on the participant to participate should be considered. Might there be costs associated with participating that might put more burden on specific individuals?

> *Example:* Alma is seeking treatment for depression but has carer responsibilities that make it difficult for her to attend regularly scheduled talking therapy sessions or exercise classes. She may be better able to pick up a prescription from a pharmacy, but struggles to find time to review medication adjustments with her provider, and her pharmacy is also quite far away. Enabling Alma to adhere to an intervention might be best achieved by providing back-up caring support, so that she is able to regularly access the intervention.

Continued access: Different types of intervention also vary in whether and how participants can continue to access them after the intervention ends. For instance, with medication-based interventions, access ends when the prescription finishes. For talking therapies such as cognitive–behavioural therapy (CBT), participants ideally can continue to use the tools they have learned beyond the end of the sessions, but direct access to the therapist ends once the course of treatment is finished. With exercise interventions, there is more likelihood for participants to be able to continue to use the resource autonomously, which could lead to lifestyle and/or identity-based adjustments.

> *Example:* Luca is offered a referral to a running group as part of his treatment plan for depression. Although he and his provider initially intend him to participate for just 6 weeks, he finds the group atmosphere friendly and informative. Following the 6 weeks, Luca pauses running briefly but, encouraged both by his own progress as a runner and messages from the other group members wondering when he is coming back, he decides to return to the group. Some 6 months later, Luca is training for a 10 k charity run and has formed meaningful friendships with several members of the group.

Cultural considerations: Different interventions are more acceptable to different groups of people based on characteristics of the intervention. Some interventions are associated with more stigma than others, and which stigma for which type of intervention (medication, counselling, exercise) can vary across populations and cultures (Qin 2020). Offering culturally oriented exercise interventions could make them more relevant, inclusive and accessible (Delattre 2024).

> *Example:* Sara has been offered the opportunity to participate in salsa dance classes as part of her course of treatment for depression, and is potentially interested, as she has had some uncomfortable conversations with family members about the idea of her doing talking therapy. However, the salsa classes are open to everyone (mixed gender), and her religion prevents her from participating in an exercise intervention of this nature with men present. She opts out of taking part and asks for the providers to make single-gender classes available in the future.

Personalisation considerations: Offering more personalisation in interventions, rather than 'one-size-fits-all', is increasingly thought to be important and affects the accessibility of an intervention. For instance, there is a need for more effective and accessible treatments for young people with depression; interventions that they perceive as generic, versus designed with them in mind, may not meet their needs, which can have an impact on treatment-seeking, adherence and efficacy (Barnett 2021). Additionally, consideration should be given to who is able to take part and what modifications need to be made for differently abled individuals.

> *Example:* Sam hopes that being more active will be a valuable part of treating their depression. However, their chronic arthritis makes many traditional forms of exercise uncomfortable. Sam is referred to a low-impact swimming club where a trained facilitator offered a seated exercise intervention with modifications of the movements and frequent breaks.

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

with pharmacological, psychological and 'other' treatments. The review included 73 randomised control trials (RCTs) (≥3624 participants completing the trials), with 69 contributing to the meta-analysis. Exercise was associated with a moderate reduction in post-treatment depressive symptoms compared with control and showed post-treatment effects comparable to those of psychological and pharmacological interventions. However, certainty of evidence ranged from low to moderate, and evidence for longer-term outcomes remained (very) uncertain, with only few trials reporting on this.

Method

Who was included?

The review included adult participants (≥18 years) with depression, as determined by any diagnostic method (including psychiatric interview and depression scales), in any setting (including inpatient), regardless of severity or comorbidities. The depression scales identified were: the Beck Depression Inventory (BDI); Hamilton Rating Scale for Depression (HRSD); Patient Health Questionnaire 9 (PHQ-9); Minnesota Multiphasic Personality Inventory (MMPI); Geriatric Depression Scale (GDS); Center for Epidemiologic Studies Depression scale (CES-D); Cornell Scale for Depression in Dementia (CSDD); and Chinese Classification and Diagnostic Criteria for Mental Health Disorders. Individuals with dysthymia, postnatal depression, anxiety or neurotic disorders were not included. This focus on only individuals with diagnosed depression is a clear methodological strength, as it ensures comparability across RCTs. At the same time, allowing variation in severity, comorbidity and treatment setting increases the generalisability of the findings to routine clinical practice.

What did the intervention look like?

The intervention included any type of exercise defined as 'planned, structured and repetitive body movements done to improve or maintain one or more components of physical fitness' (American College of Sports Medicine 2001; ACSM), provided for at least a week. Using a strict, established definition ensures that highly similar interventions are grouped together, making comparisons meaningful and improving the ability to replicate findings in future studies, and minimises ambiguity. For instance, exercise types like tai chi, qigong, yoga, pilates and stretching were not included as interventions in this review, and therefore it must be kept in mind that any effects reported by the review do not apply to those types of exercise. Subgroup analyses (i.e. analyses asking the same research question in a subset of studies/participants) examined whether outcomes differed by exercise type (aerobic, resistance or mixed), intensity (very light, light, moderate or vigorous) or number of sessions (0–12, 13–24, 25–36 or 37+).

What was exercise compared with?

First, exercise was compared with several subgroups of control conditions, including: (a) placebo; (b) no treatment, waiting list control, usual care and self-monitoring; (c) comparison of an established treatment (e.g. CBT) with or without exercise; (d) stretching, meditation, relaxation; and (e) social/casual conversation. Second, exercise was compared with pharmacological, psychological and other (occupational, art or bright light therapy; supplements; reflexology; and health education) interventions aimed at improving mood. Combination treatments including exercise were not included as it would be very difficult to separate the effects of each component. This commentary focuses on comparisons with control, psychological and pharmacological interventions, as we deemed these most clinically relevant.

The use of different types of control comparator might influence estimated treatment effects (Box 2). This is complicated by some included studies not reporting clearly what other active interventions (e.g. concomitant treatments) might be received alongside exercise (both as usual care or as part of the intervention or comparator groups), as reported in the review limitations. Additionally, although the comparison of established treatment plus exercise versus established treatment alone is a welcome addition to the review in light of concerns of potential additive effects, its grouping as a control comparison may not be appropriate (Box 2). The review's comparison between exercise and control may be affected by these choices, potentially yielding artificially reduced effect sizes that imply that the effects of exercise are more limited than they actually are.

What outcomes were assessed?

The primary outcome was change in depressive symptoms directly following intervention and at long-term follow-up (if reported), measured using various depression scales (listed above). Secondary outcomes included quality of life, cost, adverse events, acceptability of treatment (proxied by attendance and completion rate) and generalisability (proxied by number screened and recruited). This commentary focuses on the primary outcome, as we deemed this most clinically relevant.

Database search

The review authors searched five electronic databases, examining publications from February 2010

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

BOX 2 General considerations for comparators

In a meta-analysis, the focus is often comparing the effects of two conditions on an outcome measure along the same time course. Cochrane Reviews frequently include multiple comparisons against the main intervention, including the following.

- Control – includes groups with no intervention, on a waiting list, accessing usual care (which might involve pharmacological, psychological or other interventions) or accessing a standard therapy/intervention that will also be received by the experimental intervention group.
- Placebo – includes groups receiving an intervention that resembles a treatment but is not. There are different types of placebo, including standard and active placebo. For example, one group receives an antidepressant medication and the other a sucrose pill.
- Other active interventions – includes groups that receive a variation of the same intervention (e.g. different doses or lengths of treatment), or a different intervention (e.g. two types of antidepressant).

For a more detailed overview of different types of controls and placebos in mental health research, see Faltinsen et al (2022).

The comparisons against placebo and different types of control comparators focus on assessing the effectiveness of an intervention and are often analysed jointly. However, different types of control comparators might yield different outcomes for experimental interventions in trials and influence meta-analysis results and variability. A Cochrane methodological review about control interventions in randomised trials involving people with mental disorders found that trials using placebo comparators may estimate smaller experimental intervention effects than trials using waiting list or no-treatment comparators (Faltinsen 2022). This can be better understood when considering that the experience of someone receiving no intervention is likely different from the experience of someone receiving a placebo with effective blinding, influencing their expectation of effect and thereby the difference in reported mental health outcomes. That is, someone who is knowingly receiving no treatment may be less likely to expect any changes in symptoms compared with someone receiving an effectively blinded placebo, who may believe they are receiving an active treatment and may therefore expect to experience beneficial effects. This positive expectation may already provide beneficial mental health effects.

It can also be debated whether the comparison of established treatment plus the intervention versus established treatment alone should be considered an intervention versus control comparison, as was done in this review (Clegg 2026). Although at first glance it may be thought to reflect a pure intervention effect ((A + B) − B = A), it is actually a bit more complicated than that. Namely, this comparison specifically informs us of the *combined* benefit of the two treatments rather than the effect of the intervention of interest alone ((A + B) − B = A+), as is the case for other control conditions. Therefore it may be better assessed as a separate comparison rather than a control comparison.

As illustrated in this box, the types of comparator included as control conditions can have a considerable impact on reported estimates of intervention effects. The field of mental health research, at large, should consider developing standardised methodological guidelines for the design, inclusion and reporting of control interventions.

to November 2023 for four of these, and only publications from February 2010 to February 2016 for the fifth. They additionally searched three electronic registries to identify relevant ongoing trials, most recently in September 2025. This yielded 73 trials in total, with 35 new trials added since the previous version of the review.

Two authors independently assessed the risk of bias (Box 3) (Higgins 2011), which was subsequently checked by all authors working in pairs. Of note, all 73 trials were scored at 'high risk of bias' for the blinding (masking) of those delivering and receiving the intervention. This is unavoidable for exercise interventions, and although it is important to report these blinding issues in light of risk of bias, we encourage the reader to consider how this can negatively affect certainty of evidence and thereby trust in the reported findings (Box 3).

Results

Is exercise better than control at reducing depressive symptoms?

Compared with control, exercise moderately reduces depressive symptoms post-treatment (standardised mean difference s.m.d. = −0.67, 95% CI −0.82 to −0.52; $I^2 = 61\%$, $P < 0.001$; 57 trials, 2189 participants) and at long-term follow-up (s.m.d. = −0.53, 95% CI −1.11 to 0.06; $I^2 = 87\%$, $P < 0.001$; 9 trials, 405 participants). However, these findings are of low and very low certainty of evidence respectively. This is the result of unclear or high risk of bias in various domains encompassing a number of sources (Box 3: sources of bias 1, 2, 6, 7), substantial to considerable heterogeneity and a confidence interval with a lower limit that is not clinically meaningful (Box 4). Interestingly, sensitivity analyses with lower risk of bias also indicated a moderate, though slightly smaller, effect of exercise relative to control on post-treatment depressive symptoms (s.m.d. = −0.46, 95% CI −0.88 to −0.04; $I^2 = 78\%$, $P < 0.001$; 7 trials, 447 participants).

Subgroup analyses suggest that this effect of exercise applies to all exercise types, intensities and number of sessions, with no significant subgroup differences (type: $P = 0.60$; intensity: $P = 0.72$; sessions: $P = 0.32$), and relative to all types of control ($P = 0.13$) (Box 2).

Is exercise better than established treatments at reducing depressive symptoms?

Exercise may have a similar effect to psychological treatment on depressive symptoms post-treatment (s.m.d. = 0.03, 95% CI −0.16 to 0.23; $I^2 = 0\%$,

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

BOX 3 Risk of bias

The 'risk of bias' is calculated for randomised controlled trials (RCTs) to indicate the risk of underestimating or overestimating the actual effect of the intervention. This can happen as a result of, for instance, an imperfect study design (such as due to oversight or lack of resources), limited methodological rigour, or issues with analysing or reporting findings. It is important to note this risk, as it has implications for how we can interpret trial results: that is, a high risk of bias limits our ability to trust that the RCT's findings accurately reflect real-world effects.

For meta-analyses, risk of bias in individual trials is incorporated in assessing the 'certainty of evidence' of meta-analytic findings. Certainty of evidence is, as the name suggests, a metric that indicates how certain we can be that the effect estimated in a meta-analysis is actually close to the true effect and, subsequently, whether we should use it to inform real-world, clinical decision-making. It is commonly assessed in Cochrane Reviews using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework (Warner & Raslescu (2025) give a detailed explanation). If many RCTs included in a meta-analysis are considered at high risk of bias, this will negatively affect certainty of evidence of the meta-analysis, limiting confidence in the reported effects and thereby restricting our ability to provide recommendations for clinical practice based on these meta-analytic findings.

Different tools can be used to determine risk of bias. The authors of this review (Clegg 2026) used the Cochrane Collaboration risk of bias tool RoB 1 (Higgins 2011), which considers the following sources of bias for each RCT.

(1) Random sequence generation – participants should be randomly allocated to a specific intervention for the greatest likelihood that comparable groups are created. This is important as it has implications for outcome interpretations. For example, if an RCT is trying to ascertain how a new type of antidepressant medication affects depressive symptomatology, compared with placebo, and the two groups differ significantly in average depression severity, this could greatly affect how results can be interpreted. It becomes unclear whether any observed differences are due to differences in intervention (medication versus placebo) or due to group characteristics (mild versus severe depression, on average). Although randomisation does not ensure that groups are identical, it ensures that any differences are due to chance and not due to researcher selection, as otherwise researchers may inadvertently bias allocation.
(2) Allocation sequence concealment – adequate measures should be taken to conceal method of allocation to the interventions, so that experimenters cannot foresee which intervention a participant will receive when enrolling them in the trial. This is important because knowing allocations could lead researchers to selectively include participants in specific interventions. For instance, experimenters may be more inclined to enrol someone with severe symptoms into the active treatment condition rather than the placebo condition, either consciously or subconsciously. It is important to prevent this, for instance by having someone not involved with data collection to prepare pill packages (medication or placebo) and not sharing which participant numbers receive which condition.
(3) Blinding of participants: adequate measures should be taken to prevent participants from knowing which intervention they are receiving, as it may affect their engagement with and belief in the intervention, which could interfere with outcome measures.
(4) Blinding of those delivering the intervention: adequate measures should be taken to prevent the people delivering the intervention from knowing which intervention they are providing, as it may affect how they do so and/or how they interact with participants. For instance, they may present with more enthusiasm and confidence if they are giving a treatment they believe will be beneficial relative to one they think has no merit (e.g. placebo).
(5) Blinding of outcome assessors: adequate measures should be taken to prevent the people who assess and report the outcome measures from knowing which intervention was provided, as such knowledge may make them more likely to interpret and thus report findings in a biased way. For example, a psychiatrist who is assessing a participant's symptoms as a research outcome should not know the participant's group allocation (intervention versus placebo); the participant who self-reports on a task outcome should not know their group allocation; and the researcher who looks at the data preliminarily should not know whether the data comes from the intervention or placebo group when deciding which outlier data points to remove.
(6) Incomplete outcome data: adequate measures should be taken to prevent incomplete outcome data from introducing bias, as some outcome data may be missing for a reason relevant to the trial. For instance, if individuals with more severe depression are more likely to drop out of an antidepressant trial owing to a lack of motivation, they will be underrepresented in the outcome data because of something inherently linked to the trial's research question. Such bias is best prevented through use of intention-to-treat (ITT) analyses, if appropriate. ITT analyses include all randomised participants, within their allocated group, regardless of whether they received and completed the planned intervention in the correct manner. In other words, anyone who was intended to receive an intervention, even if they did not complete the study, is included in such analyses and grouped by the intended intervention (or comparator) group.
(7) Selective reporting: all results should be reported, regardless of whether they are significant, as selective reporting can skew results and their interpretations. For a detailed discussion of selective reporting and publication bias, see Norris et al (2012). In addition, findings should be reported with adequate detail to facilitate inclusion in meta-analyses.
(8) Other potential biases.

Together, these eight sources of bias are grouped into six domains: risk of selection bias (sources 1, 2), performance bias (3, 4, 5), detection bias (3, 4, 5), attrition bias (6), reporting bias (7) and any other bias (8). All included trials are rated as having a 'low', 'high' or 'unclear' risk of bias for these eight sources. For a detailed description of how this was decided by Clegg et al (2026), and how each trial was rated, please see the full Cochrane Review.

There are some limitations regarding blinding that should be considered. First, even when adequate measures are taken to blind a participant, accidental unblinding may occur. For instance, participants may be clued in to which condition they have been allocated if they experience a specific side-effect. Effort should be made to record and report this and other accidental unblinding scenarios. Second, and particularly relevant to this review, adequate blinding may be inherently impossible for some interventions. For instance, physical interventions like exercise cannot adequately blind individuals providing and receiving them, leading to a high risk of bias in the performance and detection bias domains. This is a common and understandable limitation, and it has been debated whether these two domains should be included in risk of bias assessments for trials using such interventions (see e.g. Kamper 2018). After all, if such trials can never achieve a rating beyond 'low' for these domains, their reported risk of bias will consistently be higher, leading to lower reported certainty of evidence and less trust in reported findings.

Meta-analyses may include sensitivity analyses to ascertain whether any particular bias skewed outcomes. For instance, in the current review (Clegg 2026), sensitivity analyses were used to assess whether the effect of exercise relative to control on post-treatment depressive symptoms would have been different if the meta-analysis had only included studies at low risk of bias for (a) allocation concealment, (b) blinding of outcome assessor, (c) incomplete outcome data (through use of ITT analyses) or (d) all three.

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

BOX 4 Heterogeneity, confidence intervals, imprecision and inconsistency

In addition to high or unclear risk of bias, certainty of evidence can also be decreased by inconsistency, indirectness, imprecision and publication bias (Warner 2025). This box briefly describes the concepts of heterogeneity and confidence intervals, and how they are linked to imprecision and inconsistency and, thereby, certainty of evidence in meta-analyses.

Heterogeneity: refers to variability between studies, for instance as a result of differences in study design (e.g. intervention duration/dose), population studied (e.g. average depression severity) and/or outcome measurement tools used (e.g. clinical diagnosis, Beck Depression Inventory (BDI); Hamilton Rating Scale for Depression (HRSD)). Such variability can affect the reported intervention effect, changing its size or even its direction. For instance, a trial providing a week-long exercise intervention of 30 min per day may have a more limited effect than a trial providing a week-long exercise intervention of 1 h per day. Additionally, a different intervention effect may be reported when the intervention is provided to individuals with severe depression as opposed to individuals with milder depressive symptoms. Thus, heterogeneity can lead to more inconsistency across RCT findings, and it makes meta-analytic estimates of intervention effects less precise, reducing certainty of evidence.

Heterogeneity can be assessed using the following statistics (both reported in Clegg et al's (2026) review):

(a) the chi-squared (χ^2) test: assesses whether there is any heterogeneity at all, with $P < 0.10$ generally taken to indicate the presence of significant heterogeneity;
(b) the I^2 statistic: assesses the extent of heterogeneity, ranging from 0 to 100% (0–40%, might not be important; 30–60%, moderate; 50–80%, substantial; and 75–100%, considerable).

Standardised mean difference (s.m.d.): indicates the estimated effect of an intervention (relative to a comparator) on a particular outcome variable (e.g. depressive symptom severity) that was measured with different measurement tools across trials (BDI, HRSD, etc.). Generally, an effect estimate is considered small (s.m.d. = 0.2), moderate (s.m.d. = 0.5) or large (s.m.d. = 0.8). In addition to this theoretical classification, it is important to consider whether an effect is meaningful in clinical practice. This can be determined using a pre-established minimal clinically important difference (MCID). In the current review, Clegg et al (2026) identified this as a score reduction of five points, equivalent to an s.m.d. of 0.45, for the BDI.

Confidence intervals (**CIs**): reported alongside the estimated effect of the intervention of interest, these give a range within which we can reasonably assume to find the intervention's actual, real-world effect (i.e. true effect). Different confidence levels can be used, with a 95% confidence level reported most often. In simple terms, a 95% CI can be said to indicate the range within which we can be 95% sure to find the true effect of the intervention. The width of the CI as well as the values encompassed within it affect our certainty of the estimated effect and, thereby, how we interpret the reported finding and its implications for clinical practice. Whereas a narrow CI indicates a precise estimate, a wide CI indicates imprecision and therefore more uncertainty regarding the estimated effect. Next to this, the lower and upper limits of the CI are important to take into consideration as well. For instance, consider the following fictitious meta-analytic findings regarding the effect of psychotherapy versus waiting list control on depressive symptoms.

(a) s.m.d. = −0.65 (95% CI −0.80 to −0.50): the estimated effect indicates that psychotherapy may moderately reduce depressive symptoms compared with waiting list control. The CI is narrow, indicating precision, and both the upper and lower limit of the CI suggest that psychotherapy may reduce depressive symptoms (from a moderate to a large extent). Therefore, these findings support the use of psychotherapy in the treatment of depression.
(b) s.m.d. = −0.20 (95% CI −0.40 to 0): the estimated effect indicates that psychotherapy may lead to a small reduction in depressive symptoms compared with waiting list control. The CI is equally as narrow, as in the previous example, again indicating precision. However, the values included in the CI now tell a different story: psychotherapy relative to waiting list control may reduce depressive symptoms *or* may not affect depressive symptoms differently. Therefore, certainty of evidence is lowered, and these findings provide only tentative support for the use of psychotherapy in the treatment of depression.
(c) s.m.d. = −0.65 (95% CI −1 to −0.3): the estimated effect again indicates that psychotherapy may reduce depressive symptoms relative to waiting list control. The CI is wide, indicating imprecision. Both the lower and upper limit of the CI show a negative value, thus indicating that we can reasonably expect that psychotherapy reduces depressive symptoms compared to waiting list control. However, we are less certain to what extent, with the CI ranging from values denoting a small to large, and not clinically meaningful to very clinically meaningful, effect. Therefore, certainty of evidence is lowered, and these findings provide only limited support for the use of psychotherapy in the treatment of depression.
(d) s.m.d. = −0.6 (95% CI −1.2 to 0.3): although the estimated effect indicates that psychotherapy may reduce depressive symptoms moderately compared with waiting list control, the indicated CI reduces our certainty regarding this finding. Namely, the CI is wide, indicating imprecision. Moreover, the values of the CI lower and upper limits indicate that SSRIs may have vastly different effects on depressive symptoms, anywhere between a large reduction (−1.2) to a small increase (0.3).

As can be understood from these examples, the implication of reported findings depends heavily on the CI – both its width and the values of its upper and lower limits. The width of CIs can be affected by various factors. For instance, for individual RCTs, a smaller sample size leads to wider CIs, denoting a higher degree of imprecision. For meta-analyses, CIs widen as a result of (a) fewer included RCTs, (b) less precise RCTs (e.g. because of their small sample size), (c) high heterogeneity among the different RCTs and (e) use of a random effects model. Thus, a limited evidence base (i.e. few RCTs available, with limited sample sizes) will negatively affect confidence intervals, particularly if there is also a lot of heterogeneity.

$P = 0.85$; 10 trials, 414 participants) and at long-term follow-up (s.m.d. = −0.11, 95% CI −0.48 to 0.26; $I^2 = 0\%$, $P = 0.47$; 4 trials, 114 participants). Certainty of evidence is moderate for the short-term effect, but low for the long-term effect because of unclear or high risk of bias in various domains (bias sources 1, 2, 6, 7; Box 3) and imprecision (Box 4), limiting confidence in the reported long-term effect.

Additionally, exercise may affect post-treatment depressive symptoms similarly to pharmacological treatment (s.m.d. = −0.11, 95% CI −0.33 to 0.10; $I^2 = 0\%$, $P = 0.69$; 5 trials, 330 participants).

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

Certainty of evidence for this is low owing to unclear or high risk of bias in various domains (bias sources 1, 2, 6, 7; Box 3) and because the lower limit of the confidence interval does not reach clinical significance (Box 4). Contrastingly, depressive symptoms at longer-term follow-up may be slightly reduced following exercise compared with pharmacological treatments (mean difference = −0.40, 95% CI −0.80 to 0.00; 1 trial, 58 participants). However, the evidence is very uncertain as it comes from only a single trial with a very specific population (i.e. community-dwelling older adults) and is at unclear or high risk of bias in various domains (bias sources 1, 2, 6, 7; Box 3), and because the lower limit of the confidence interval does not signify a clinically meaningful effect (Box 4).

Discussion

What does this mean for my patient?

Clegg et al's (2026) review findings suggest that exercise meeting the ACSM definition, delivered for at least a week, may be useful for managing depressive symptoms, particularly for patients who are not receiving any other active treatments. This applies across exercise type (not including e.g. yoga/stretching), intensity and number of sessions, and is in line with a previous network meta-analysis using a wider sample and definition of exercise (Noetel 2024). Direct comparisons also suggested antidepressant effects of a similar magnitude to pharmacological and psychological treatments.

However, certainty of evidence is currently limited by heterogeneity, imprecision and a limited evidence base, and longer-term effects remain unclear as they are underexplored. Therefore, based on this review, exercise should not currently be viewed as a universal replacement for established pharmacological or psychological treatments. However, clinicians may consider recommending structured exercise interventions to patients with depression on waiting lists, not receiving treatment, or less likely to adhere or respond well to conventional treatments.

The research: what comes next?

As noted by the review authors, future studies into exercise for depression must include larger sample sizes and improved quality and methodological rigour, to reduce heterogeneity and risk of bias. We additionally urge the careful selection of control comparators in future trials (Box 2), use of follow-up measures of depressive symptomatology, and an investigation of the effects of different exercise types, sessions and intensities to inform individualised care (Box 1). These factors warrant consideration, with previous meta-analytic findings tentatively suggesting a dose–response curve for exercise intensity and differential effects for different exercise modalities (moderated by participant age and gender, and social component) (Noetel 2024). Future trials should also include participants from a wide variety of backgrounds, to ensure generalisability of findings, and health equity and treatment cost should be explored.

To inform individualised prescribing and aid accessibility, we recommend that future review updates, using higher-quality individual trials, assess how differences in baseline depression severity and intervention aspects such as delivery method (trained instructors versus self-directed), social connection opportunities (group-based versus solo; Delattre 2024), and setting (indoor versus outdoor) affect exercise's antidepressant effects. Furthermore, we recommend examining the merits of exercise as an adjunct treatment through a main comparison (Box 2). Exercise has been reported to provide the greatest benefit combined with other treatments (Gonzales-Garcia 2026) and can easily be implemented alongside other intervention types, underscoring its potential as an adjunct treatment.

Finally, using higher-quality RCTs, a separate review should explore how physical activity interventions not included in the current review (e.g. yoga, tai chi) could benefit depressed individuals as compared with control and other treatments, considering that these exercise types may be more accessible to patients with disabilities (Box 1) and have tentatively shown promising antidepressant effects (Noetel 2024).

Conclusion

This review suggests that exercise (as defined here) reduces depressive symptoms post-treatment compared with control and may provide benefits comparable with psychological and pharmacological treatments, although this must be interpreted with caution considering the uncertainty of the findings. Long-term effects remain underexplored. However, the effects of exercise on depression are tentatively promising and merit further investigation, particularly from a personalised medicine point of view, to reliably inform clinical practice and contribute to the range of available evidence-based approaches to mitigating depression.

Acknowledgements

Artificial Intelligence use: The authors made use of an AI-assisted language tool (ChatGPT Edu, based on GPT-5 architecture), accessed in June 2026 without modification, to support language editing of

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press

the draft manuscript. All outputs were reviewed, verified and substantially revised by the authors, who take full responsibility for the content.

Author contributions

This article was created by all three authors. L.L.M.B. wrote the first and second draft, apart from Boxes 1 (created by B.D.) and 2 (created by D.A.B.). All authors worked together to revise these two drafts, and subsequent drafts were devised together.

Funding

This research was funded by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre: Oxford Health. The views expressed are those of the authors and not necessarily those of the NIHR, the Department of Health and Social Care, or the NHS trusts that the authors are affiliated to.

Declaration of interest

None.

References

American College of Sports Medicine (2001) *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription* (4th edn). Lippincott Williams & Wilkins.

American Psychiatric Association (2013) Major depressive disorder. In *Diagnostic and Statistical Manual of Mental Disorders* (5th edn) (DSM-5): 160–8. APA Publishing.

Barnett P, Arundell L-L, Matthews H, et al (2021) 'Five hours to sort out your life': qualitative study of the experiences of university students who access mental health support. *BJPsych Open*, **7**: e118.

Clegg AJ, Hill JE, Mullin DS, et al (2026) Exercise for depression. *Cochrane Database of Systematic Reviews*, **1**: CD004366 (https://doi.org/10.1002/14651858.CD004366.pub7).

Cuijpers P, Stringaris A, Wolpert M (2020) Treatment outcomes for depression: challenges and opportunities. *Lancet Psychiatry*, **7**: 925–7.

Delattre B, Applewhite B, Buckman JEJ, et al (2024) Social dance and movement for mental health: a narrative review. *Mental Health Science*, **2**: e62.

Fabiano N, Puder D, Stubbs B (2025) Could not prescribing exercise for depression be psychiatric malpractice? *British Journal of Sports Medicine*, **59**: 1388–9.

Faltinsen, E, Todorovac, A, Bruun, LS, et al (2022) Control interventions in randomised trials among people with mental health disorders. *Cochrane Database of Systematic Reviews*, **4**: MR000050.

Gonzalez-Garcia X, Garcia A, García-Pazo P, et al (2026) Lifestyle interventions as adjuvant treatments for depression: an umbrella review of systematic reviews and meta-analyses. *Complementary Therapies in Medicine*, **98**: 103336.

Henking C, Reeves A, Chrisinger B (2023) Global inequalities in mental health problems: understanding the predictors of lifetime prevalence, treatment utilisation and perceived helpfulness across 111 countries. *Preventive Medicine*, **177**: 107769.

Higgins JPT, Altman DG, Gotzsche PC, et al (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, **343**: d5928.

Hird, EJ, Slanina-Davies, A, Lewis, G, et al (2024) From movement to motivation: a proposed framework to understand the antidepressant effect of exercise. *Translational Psychiatry*, **14**: 273.

Kamper SJ (2018) Blinding: linking evidence to practice. *Journal of Orthopaedic and Sports Physical Therapy*, **48**: 825–6.

Marx W, Penninx BWJH, Solmi M, et al (2023) Major depressive disorder. *Nature Reviews Disease Primers*, **9**: 44.

Meng Y, Chiu C, Kapoor M, et al (2024) Patient perceived barriers and enablers to medication adherence in the treatment of depression: a qualitative study. *Journal of Primary Care & Community Health*, **15**: 21501319241286313.

Noetel M, Sanders T, Gallardo-Gómez D, et al (2024) Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials. *BMJ*, **384**: e075847.

Norris SL, Moher D, Reeves BC, et al (2012) Issues relating to selective reporting when including non-randomized studies in systematic reviews on the effects of healthcare interventions. *Research Synthesis Methods*, **4**: 36–47.

O'Neill, J, Tabish, H, Welch, V, et al (2013) Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *Journal of Clinical Epidemiology*, **67**: 56–64.

Qin X, Hsieh C-R (2020) Understanding and addressing the treatment gap in mental healthcare: economic perspectives and evidence from China. *Inquiry*, **57**: 0046958020950566.

Ross RE, Van Derwerker CJ, Saladin ME, et al (2023) The role of exercise in the treatment of depression: biological underpinnings and clinical outcomes. *Molecular Psychiatry*, **28**: 298–328.

Rush AJ, Warden D, Wisniewski SR, et al (2009) STAR*D revising conventional wisdom. *CNS Drugs*, **23**: 627–47.

Warner G, Raslescu A (2025) Treating post-traumatic stress disorder online: uncertain evidence for internet-delivered cognitive and behavioural therapies. *BJPsych Advances*, **32**: 152–8.

Welch VA, Petkovic J, Jull J, et al (2024) Equity and specific populations [last updated Oct 2019]. In *Cochrane Handbook for Systematic Reviews of Interventions Version 6.5* (eds JPT Higgins, J Thomas, J Chandler, et al): ch 16. Cochrane (https://www.cochrane.org/handbook).

Welkerling J, Niess A, Schneeweiss P, et al (2026) Single bout of exercise reduces self-reported and decoded rumination in favor of distraction in patients with major depression. *Journal of Affective Disorders*, **397**: 120829.

Wittmann L, Blomert E, Linden M (2023) Patients' perception of side effects in cognitive-behavior, psychodynamic, and psychoanalytic outpatient psychotherapy. *Psychotherapy Research*, **35**: 112–24.

World Health Organization (2026) *Depressive Disorder (Depression) [last updated 11 Sep 2026]*. WHO (https://www.who.int/news-room/fact-sheets/detail/depression).

https://doi.org/10.1192/bja.2026.10243 Published online by Cambridge University Press