

This is a repository copy of *A systematic review and network meta-analysis of psychological, psychosocial, pharmacological, physical and combined treatments for adults with a new episode of depression.*

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/215658/>

Version: Published Version

Article:

Mavranetzouli, Ifigeneia, Megnin-Viggars, Odette, Pedder, Hugo et al. (14 more authors) (2024) A systematic review and network meta-analysis of psychological, psychosocial, pharmacological, physical and combined treatments for adults with a new episode of depression. *EClinicalMedicine*. 102780. ISSN 2589-5370

<https://doi.org/10.1016/j.eclinm.2024.102780>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

A systematic review and network meta-analysis of psychological, psychosocial, pharmacological, physical and combined treatments for adults with a new episode of depression



Ifigenia Mavranzeouli,^{a,*} Odette Megnin-Viggars,^a Hugo Pedder,^b Nicky J. Welton,^b Sofia Dias,^c Edward Watkins,^d Neil Nixon,^e Caitlin H. Daly,^b Edna Keeney,^b Hilary Eadon,^f Deborah M. Caldwell,^b Katriona J. M. O'Donoghue,^f Sarah Stockton,^g Stephanie Arnold,^f James Thomas,^h Navneet Kapur,^{i,j,k} and Stephen Pilling^{a,k}



^aCentre for Outcomes Research and Effectiveness, Research Department of Clinical, Educational & Health Psychology, University College London, London, UK

^bPopulation Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

^cCentre for Reviews and Dissemination, University of York, York, UK

^dFaculty of Health and Life Sciences, University of Exeter, Exeter, UK

^eSchool of Medicine, University of Nottingham, Nottingham, UK

^fNational Institute for Health and Care Excellence, Manchester, UK

^gGateshead Health NHS Foundation Trust Library, Gateshead, UK

^hEPPI-Centre, Social Research Institute, University College London, London, UK

ⁱUniversity of Manchester, Manchester, UK

^jMersey Care NHS Foundation Trust, Liverpool, UK

Summary

Background Various effective treatments for depression exist. We aimed to identify the most effective first-line treatments for new episodes of less and more severe depression (defined by depression scale cut-off scores), to update NICE guidance on the management of Depression in Adults in England.

Methods Systematic review and network meta-analysis of randomised controlled trials (RCTs) published up to June 2020 (PROSPERO registration number CRD42019151328). We analysed interventions by class and individually. The primary efficacy outcome was depressive symptom change (expressed as standardised mean difference [SMD]). The review for this outcome was updated in November 2023.

Findings We included 676 RCTs, 105,477 participants and 63 treatment classes. For less severe depression, group cognitive/cognitive behavioural therapy (CT/CBT) class was efficacious versus treatment as usual [TAU], the reference treatment for this population [SMD -1.01 (95% Credible Interval [CrI] -1.76 ; -0.06)]. For more severe depression, efficacious classes versus pill placebo (reference treatment for this population) included combined individual CT/CBT with antidepressants [-1.18 (-2.07 ; -0.44)], individual behavioural therapies [-0.86 (-1.65 ; -0.16)], combined light therapy with antidepressants [-0.86 (-1.59 ; -0.12)], combined acupuncture with antidepressants [-0.78 (-1.12 ; -0.44)], individual CT/CBT [-0.78 (-1.42 ; -0.33)], mirtazapine [-0.35 (-0.48 ; -0.22)], serotonin and norepinephrine reuptake inhibitors [-0.32 (-0.43 ; -0.22)], tricyclic antidepressants [-0.29 (-0.50 ; -0.05)], and selective serotonin reuptake inhibitors [-0.24 (-0.32 ; -0.16)]. Additional treatments showed evidence of efficacy at the intervention level. Evidence for less and more severe depression was of low and low-to-moderate quality, respectively. In the 2023 update, group yoga and self-help without support emerged as efficacious for less severe depression. For more severe depression, combined group exercise with antidepressants emerged as efficacious, whereas combined light therapy with antidepressants failed to remain efficacious.

Interpretation Group CT/CBT (and possibly group yoga and self-help) appears efficacious in less severe depression, whereas antidepressants do not show evidence of effect. Combined antidepressants with individual CT/CBT, acupuncture and, possibly, group exercise, individual psychological therapies (behavioural therapies, CT/CBT) alone, and antidepressants alone appear efficacious in more severe depression. Quality of evidence, cost-

eClinicalMedicine
2024;75: 102780
Published Online xxx
<https://doi.org/10.1016/j.eclinm.2024.102780>

*Corresponding author. Centre for Outcomes Research and Effectiveness, Research Department of Clinical, Educational & Health Psychology, University College London, 1-19 Torrington Place, London, WC1E 7HB, UK.

E-mail address: i.mavranzeouli@ucl.ac.uk (I. Mavranzeouli).

^kEqual senior authors.

effectiveness, applicability and implementation issues also need to be considered when formulating clinical practice recommendations.

Funding National Institute for Health and Care Excellence.

Copyright © 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Depression; Psychological interventions; Antidepressants; Physical interventions; Network meta-analysis; Clinical guidelines

Research in context

Evidence before this study

Several published network meta-analyses (NMAs) have concluded that psychological, pharmacological and combined therapies are effective in the treatment of adults with depression, however none have made comparisons between specific psychological and pharmacological interventions and very few stratified populations by depressive symptom severity. We searched Embase, Emcare, MEDLINE, PsycINFO, CENTRAL and Cochrane Database of Systematic Reviews from inception to June 2020 to identify RCTs of psychological, psychosocial, pharmacological, physical and combined treatments used as first-line therapy for adults with a new episode of unipolar depression (not currently receiving treatment) and synthesised data using NMA techniques to address these gaps in evidence and update national guidance on the treatment and management of depression in adults in England, published by the National Institute for Health and Care Excellence (NICE). We undertook separate analyses for new episodes of less severe (subthreshold and mild) and more severe (moderate and severe) depression, as defined by clinical cut-off depression scale scores. The review was updated in 2023.

Added value of this study

Our NMA included 676 eligible RCTs (142 on less and 534 on more severe depression) assessing 63 treatment classes on 105,477 participants. Results differ from previous NMA findings in suggesting that (a) antidepressants do not show evidence of an effect against pill placebo in less severe

depression; (b) group CBT may be the most effective treatment for less severe depression; (c) the mode of delivery (individual versus group) appears to impact on CBT effectiveness; and (d) some psychological interventions (such as individual CBT, individual behavioural activation, non-directive counselling, computerised CBT with or without support) may be more effective than antidepressants in more severe depression. Results remained largely unchanged following a review update in 2023.

Implications of all the available evidence

The findings of this NMA should be considered alongside other contextual factors when making practice recommendations on the treatment of adults with depression. Our results (including the size, uncertainty and limitations of the evidence) informed the updated NICE national guideline on the treatment and management of depression in adults. Other factors that were considered when formulating recommendations included cost-effectiveness evidence, results of pairwise meta-analysis of outcomes at follow-up, quality of life and functioning data, qualitative evidence on the facilitators and barriers to treatment choice, side effects and withdrawal symptoms associated with antidepressants, applicability of the evidence, implementation issues with regard to current structure of psychological treatment services in England, and the need to provide a wide range of interventions to meet individual needs and support patient choice.

Introduction

Worldwide, major depression has a prevalence of 4.33% and is a leading cause of disability.¹ Depression is associated with substantial costs to the health service and the wider society, mainly due to productivity losses.^{2,3} Providing evidence-based, effective interventions targeting depression will increase service costs but is expected to lead to overall cost-savings via increased productivity.² Systematic reviews and pairwise meta-analyses have indicated that antidepressant medication and psychological interventions (e.g., cognitive behavioural therapy) are effective interventions for depression singly and in combination.⁴⁻⁶ However, there is limited

evidence on the direct comparison of psychological and pharmacological interventions with evidence assessments tending to be siloed, with separate large network meta-analyses (NMAs) of antidepressant medication^{7,8} or psychotherapy⁹⁻¹⁴ that do not fully address clinical choice. Of the many NMAs now available on the treatment for adults with depression,⁷⁻¹⁹ none has made comparisons between specific psychological and pharmacological interventions and few¹⁷⁻¹⁹ have assessed physical treatments. Furthermore, very few of the previous NMAs stratified populations by depressive symptom severity,^{13,15} a key factor in informing treatment choice.²⁰

We employed NMA techniques to examine the relative effectiveness, acceptability and tolerability of psychological, psychosocial, pharmacological, physical and combined treatments, used as first-line therapy for adults with a new episode of unipolar depression (not currently receiving treatment), stratified by symptom severity level, to address these gaps and update national guidance on the management of depression in adults in England, published by the National Institute for Health and Care Excellence (NICE).²¹

Methods

A systematic review of randomised controlled trials (RCTs) of treatments for adults with a new episode of unipolar depression was undertaken according to Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines for NMAs.²² The study protocol ([Appendix 1](#)) was registered on PROSPERO (CRD42019151328). The review focused on first-line treatments for a new depressive episode, as separate reviews were conducted to identify suitable treatments for people who did not respond to previous treatment and those with chronic depression.

Search strategy

Searches for RCTs (and systematic reviews of RCTs) of treatments for adults with depression were conducted in Embase, Emcare, MEDLINE, PsycINFO, CENTRAL and Cochrane Database of Systematic Reviews from inception, using relevant medical subject headings, free-text terms, and study type filters where appropriate ([Appendix 1](#)). The search was undertaken in May 2019 and updated in June 2020, with this date limit set to meet NICE guidance publication timelines (see also 2023 review update section below, specific to this publication). Additional search methods included checking reference lists of systematic reviews (identified through the electronic database search), and citation searches using Web-of-Science for included studies.

Selection criteria

RCTs were eligible for inclusion if $\geq 80\%$ of participants were adults not currently receiving treatment, initiated on first-line treatment for a new episode of unipolar depression (or subthreshold depressive symptoms) according to Diagnostic and Statistical Manual of Mental Disorders (DSM), International Classification of Diseases (ICD) or similar criteria, or by scoring above threshold on a validated depression scale. We excluded populations with perinatal depression, seasonal affective disorder, bipolar disorder, learning disabilities or in contact with the criminal justice system. Trials of further-line treatment, those specifically recruiting participants with depression and a physical health condition, or where $\geq 20\%$ of participants had psychotic symptoms or a co-existing personality disorder or

chronic depression were also excluded. We included only studies reporting data that could inform one or more outcomes of interest (described under 'Analysis plan and data extraction').

The review considered pharmacological, psychological, psychosocial, physical and combined treatments. Pharmacological treatments included selective serotonin reuptake inhibitors (SSRIs), serotonin-norepinephrine reuptake inhibitors (SNRIs), tricyclic antidepressants (TCAs), mirtazapine and trazodone. They were eligible if licensed in the UK and in routine clinical use as first-line treatments of depression. Psychological/psychosocial treatments included behavioural therapies, cognitive and cognitive behavioural therapies (CT/CBT), counselling, interpersonal psychotherapy (IPT), psychodynamic psychotherapies, psychoeducation, self-help with or without support, mindfulness, meditation or relaxation, and peer support. Physical treatments included acupuncture, exercise, yoga and light therapy. Combined treatments were also included (e.g., individual CT/CBT + antidepressants). Non-eligible treatments (such as drugs of no interest for decision-making, for example imipramine and bupropion, or broader classes and types of treatment rather than specific drug interventions, such as 'SSRIs', 'TCAs', and 'antidepressants'), were included in the analysis if they had formed a combined treatment with an eligible psychological or physical intervention or if they had been compared against an eligible psychological or physical intervention, in order to enhance connectivity of eligible treatments in the networks. However, non-eligible treatments were not considered as part of the decision problem (that is, they were not considered when formulating recommendations). Controls included treatment as usual (TAU), waitlist, no treatment, attention placebo, sham treatments, and pill placebo. [Appendix 1](#) provides a comprehensive description of eligible treatments, and details on the approach of considering ineligible pharmacological treatments in the analysis so as to enhance network connectivity.

Titles and abstracts of identified studies were screened by two reviewers for inclusion against predefined protocol criteria, until inter-rater reliability of $\geq 90\%$ agreement was observed. Initially 10% of references were double-screened; as inter-rater agreement was $>90\%$, the remaining references were screened by one reviewer. For studies considered relevant after title/abstract screening, the full text was acquired and checked for eligibility by both reviewers. Disagreements in study inclusion were resolved via discussion between the two reviewers, and consultation with a senior reviewer if necessary.

Analysis plan and data extraction

Separate analyses were conducted for new episodes of less and more severe depression. Baseline mean scores on validated depression scales were used to classify the

study population into one of these two levels of depressive symptom severity, using, as a starting point, default cut-off points for different depressive symptom levels, crosswalk tables of standardized depression measurements,^{23–26} and clinical judgement. Emphasis was given to PHQ-9 as this is the most widely used screening tool for depression in primary care in the UK, although there were concerns that default thresholds may overstate depressive symptom severity compared with other validated scales.^{27–29} Provisional cut-off points were further calibrated to allow a clinically meaningful distinction between the two severity levels, aligned with treatment decisions in clinical practice. Full details are provided in [Appendix 1](#). The final depression scale cut-off points used to classify populations into groups of adults with a new episode of less and more severe depression are shown in [Table 1](#). Where baseline scores were unavailable, severity was determined according to the study inclusion criteria or description (if unambiguous, e.g. ‘severe’ or ‘subthreshold’).

Due to the high number of included interventions, these were grouped into treatment classes, according to common theoretical structure or hypothesized mechanism of action (see [Appendix 1](#) for details).

Seven outcomes at treatment endpoint were analysed using NMA techniques:

- Efficacy:
 - o Standardised mean difference (SMD) of depressive symptom change from baseline (primary efficacy measure)
 - o Response (typically defined as $\geq 50\%$ improvement on a depression scale score) in those randomised
 - o Response in treatment completers (that is, in those who did not discontinue treatment early)

- o Remission (typically defined as a depression scale score below a threshold) in those randomised
- o Remission in treatment completers
- Acceptability: treatment discontinuation for any reason
- Tolerability: treatment discontinuation due to side effects from medication, in those who discontinued treatment.

Four of these outcomes (response and remission in treatment completers; discontinuation due to any reason and due to side effects) informed the model-based economic evaluation of first-line treatments for a new episode of depression undertaken to support the NICE guideline.

From each included study, we extracted: country, population, intervention details, outcome data, and potential risk of bias assessed using the Cochrane Risk-of-Bias tool.³⁰ Data extraction was double-checked by a second reviewer. Disagreements were resolved via consultation with a senior reviewer. A large NMA of antidepressants for the acute treatment of adults with depression⁷ was used as a source of data from non-English language and unpublished studies.

Statistical analysis

NMAs were conducted within a Bayesian framework using Markov Chain Monte Carlo simulation techniques implemented in OpenBUGS 3.1.2, which is a variant of WinBUGS.^{31–33} For the depressive symptom change, we pooled the SMD between treatments using a NMA model with normal likelihood and identity link function, accounting for different reporting formats between studies.³⁴ For discontinuation, response and remission we pooled log-odds ratios (LORs) between pairs of treatments using a NMA model with binomial likelihood and logit link function³⁴ ([Appendix 1](#) provides method details). We used random study effects models to capture the expected between-study variability in intervention effects. Intervention effects were modelled using random class models, assuming that intervention effects within each class are distributed around a common class mean effect with a within-class variance.

For each analysis we estimated mean relative effects (SMD or LOR, as relevant) between classes and between interventions, with 95% credible intervals (CrI). We also estimated mean ranks with 95%CrI for every class and intervention, where a rank of 1 indicates the best outcome. Although all connected eligible classes and interventions were included in the NMAs, we only considered results for classes and interventions tested on ≥ 50 participants (or ≥ 50 completers, for the completer outcomes) across RCTs in each analysis, as the minimum adequate evidence base to support recommendations. We determined ‘evidence of effect’ based on whether the 95%CrI crossed the no effect line. We reviewed relative effects against a reference

Scale	Threshold
HAMD (17-item, 21-item and 24-item)	16
MADRS (10-item)	22
PHQ-9	16
BDI-I (21-item)	22
BDI-II (21-item)	30
CES-D (20-item)	36
QIDS (16-item)	12
HADS-D (7-item)	12

BDI: Beck’s Depression Inventory; CES-D: Center for Epidemiologic Studies Depression Scale; HADS-D: Hospital Anxiety and Depression Scale—Depression sub-scale; HAMD: Hamilton Depression Rating Scale; QIDS: Quick Inventory of Depressive Symptomatology; MADRS: Montgomery–Åsberg Depression Rating Scale; PHQ-9: Patient Health Questionnaire.

Table 1: Cut-off points for depression scales, estimated using crosswalk tables of standardized depression measurements^{23–26} (with a score at the threshold point and above indicating more severe depression, that is, of at least moderate severity).

treatment, which was treatment as usual (TAU) and pill placebo for less and more severe depression, respectively, selected following consideration of their clinical significance, size of the evidence and connectivity in each population. TAU was relatively ill-defined in the included studies, with often little more detail than 'usual care'. However, where a study did report the proportion of participants who received different treatments within the TAU arm, and if at least 80% of participants in that arm received a consistent treatment class (for example SSRIs), then the arm was coded as such rather than as TAU. As outlined in the review protocol, where a study compared 'intervention + TAU versus TAU alone' it was recoded as 'intervention versus no treatment'. Pill placebo was originally preferred for use as the reference treatment for both less and more severe depression, because it is well-defined across RCTs. However, there was either very limited connectivity or lack of evidence for pill placebo in the less severe depression networks, so that it wasn't possible to use as the reference treatment for this population. On the other hand, TAU may be a heterogeneous condition but has been widely used as a control condition across RCTs in populations with a new episode of less severe depression; for this reason, TAU was selected as the reference treatment for less severe depression. For discontinuation due to side effects from medication, the reference treatment was pill placebo for both severity levels. We also made comparisons between active treatments at class and intervention level.

Transitivity and inconsistency checks

NMA assumes that the distribution of effect modifiers is the same across treatment comparisons ('transitivity' assumption). To control for potential effect modifiers, we aimed to reduce heterogeneity by stratifying analyses by depressive symptom severity and assigning interventions to classes using detailed definitions and considering their mode of delivery. Violations of the transitivity assumption may lead to inconsistency between direct and indirect evidence estimates.³⁵ This was assessed statistically by comparing the fit of a model assuming consistency with a model allowing for inconsistency at the intervention level, whilst still modelling class effects³⁶ (see [Appendix 1](#)). Furthermore, we compared NMA and inconsistency model effects for all comparisons and estimated differences between direct and indirect effects to identify treatment comparisons where this difference was statistically meaningful and higher than the minimal clinically important difference, determined as ± 0.5 for the SMD and ± 0.25 for log-ORs.

We further explored the transitivity between pharmacological and non-pharmacological trial participants by conducting a sensitivity analysis on the depressive symptom change outcome (SMD) that excluded pharmacological trials, and comparing the effect size and

ranking of non-pharmacological treatments between this analysis and the full analysis (which included pharmacological trials).

Quality assessment of the evidence

We evaluated the quality of the evidence with a narrative summary using the standard GRADE approach domains (risk of bias, inconsistency, publication bias, indirectness and imprecision).³⁷

Bias adjustment models

Bias adjustment models³⁸ were fitted for the outcomes of depressive symptom change (SMD), treatment discontinuation for any reason, and response in completers, testing for bias associated with small sample size³⁹ ([Appendix 1](#)). Analyses assumed possible bias in comparisons of active interventions versus inactive controls and no bias in comparisons between inactive controls or between active interventions. As an exception we assumed bias against counselling in comparisons where counselling was the control intervention, as there were concerns that non-directive counselling, when used as control, is less likely to be manual-based and delivered in a comparable number of sessions by an equivalent healthcare professional as when it is assessed as the active intervention. Magnitude of bias was assumed to be the same across all potentially biased comparisons. Where bias was indicated, results from bias-adjusted models were considered.

2023 review update

We reran the full search on 28 November 2023 and updated the NMA of the primary outcome (SMD of depressive symptom change) for less and more severe depression to explore whether there were substantial changes in conclusions between our original analysis (2020) and the updated analysis (2023). Due to the high number of records captured in the search, we first used machine learning methods to eliminate records with high likelihood of being irrelevant, before proceeding to manual selection of the remaining, potentially eligible, studies. [Appendix 2](#) provides the updated search strategy and a description of the machine learning methods employed for the review update.

Role of the funding source

This work was funded by NICE, who had no active role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Studies and treatments

The systematic search identified 3451 potentially eligible publications, of which 676 RCTs (142 on less and 534 on

more severe depression) that assessed 63 treatment classes on 105,477 participants met eligibility criteria for the NMA (Fig. 1). Participants included adults with a new episode of unipolar depression as defined according to DSM, ICD or similar criteria in 503 RCTs, or depressive symptoms as indicated by baseline depression scores on validated scales (including those with subthreshold depressive symptoms) in the remaining 173 RCTs. The median treatment duration was 6 weeks (interquartile range -IR- 6 to 8 weeks) for pharmacological interventions, 9 weeks (IR 6–12 weeks) for psychological interventions, 8 weeks (IR 6–12 weeks) for physical interventions and 12 weeks (IR 6–14 weeks) for combined interventions. Appendix 3 contains lists of included and excluded studies (with study characteristics for included studies).

The networks of depressive symptom change (SMD) included 127 RCTs, 16,829 participants, 34 classes and 76 interventions for less severe depression, and 352 RCTs, 59,350 participants, 50 classes and 99 interventions for more severe depression. See Fig. 2 for class level SMD network plots, Appendix 4 for all other network plots (class and intervention level) and information on classes/interventions, numbers of RCTs and participants on each outcome, and Appendix 5 for full

data utilised and the number of RCTs per comparison in each NMA.

Assessment of heterogeneity, inconsistency and bias

In less severe depression, heterogeneity (measured by the posterior median between-study standard deviation) was high for response outcomes, and moderate for depressive symptom change (SMD), discontinuation for any reason, and remission outcomes. In more severe depression, heterogeneity was low for remission in treatment completers and moderate for all other outcomes. In less severe depression, evidence of inconsistency was identified only for response in completers. In more severe depression, there was evidence of inconsistency in SMD, response in those randomised, and discontinuation for any reason (Appendix 6). Comparison of NMA and inconsistency model estimates identified only a small number of treatment comparisons with clinically important differences between direct and indirect effects (Appendix 7).

Evidence of small-study bias was identified for depressive symptom change (less and more severe depression), response in treatment completers (more severe depression) and treatment discontinuation for

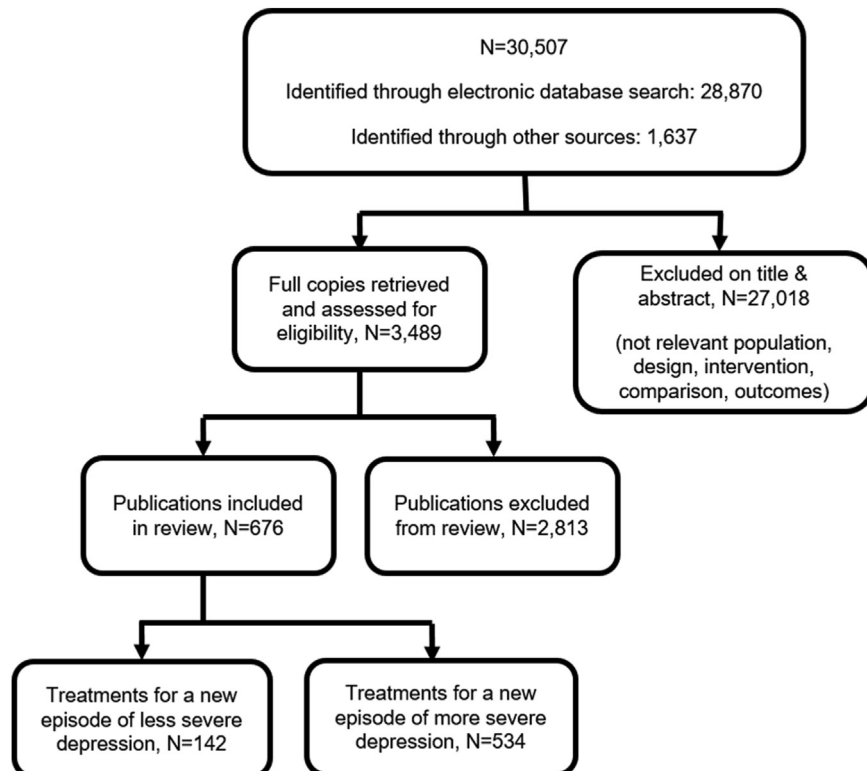


Fig. 1: Flow diagram of study selection for the systematic review and network meta-analysis.

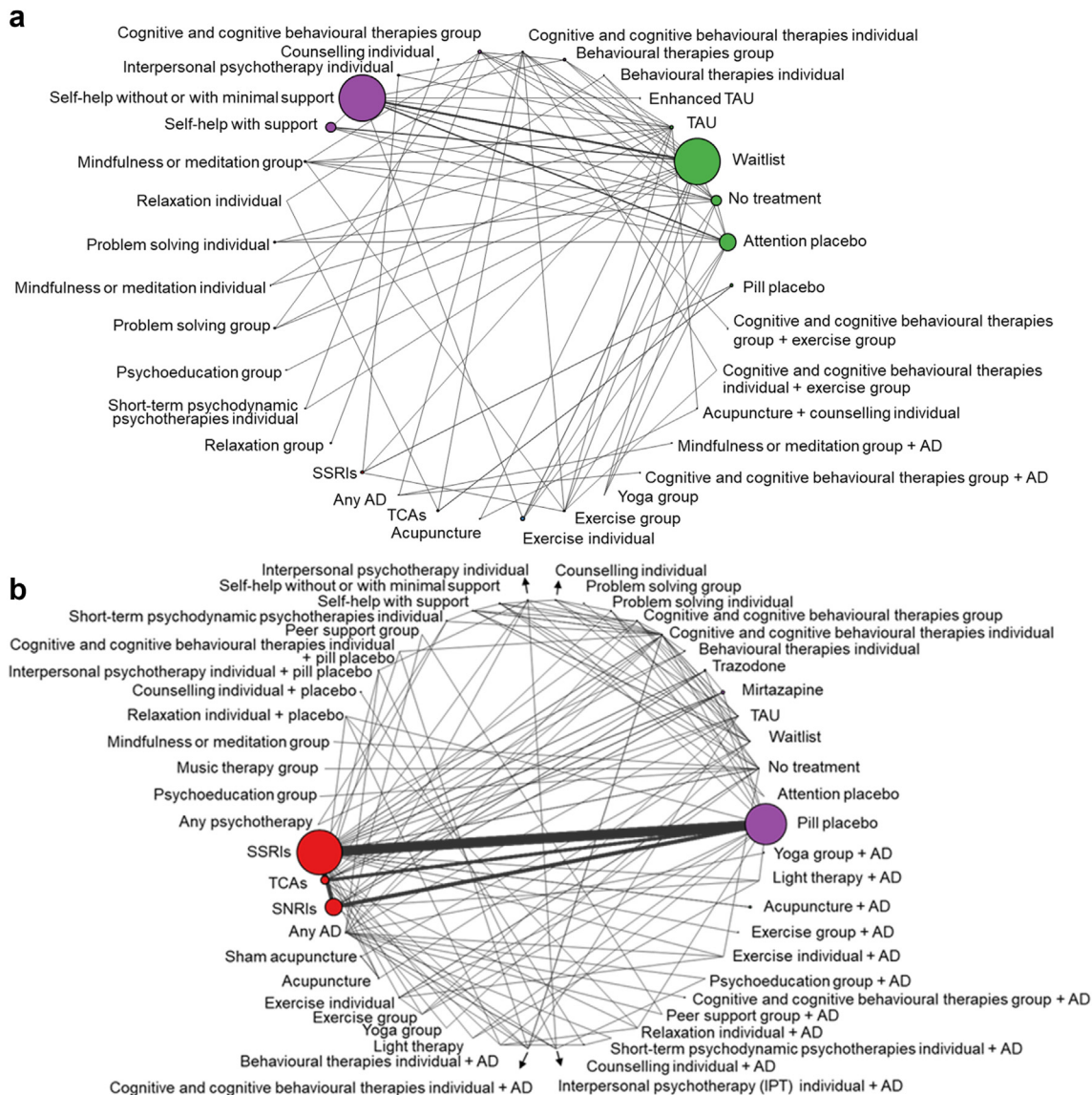


Fig. 2: Networks of the NMA of standardised mean difference (SMD) of depressive symptom change in adults with a new episode of (a) less severe depression and (b) more severe depression—treatment class level. The width of lines is proportional to the number of trials in which each direct comparison is made. The size of each circle (treatment node) is proportional to the number of participants on each treatment class across RCTs. AD: antidepressant; SNRIs: serotonin and norepinephrine reuptake inhibitors; SSRIs: selective serotonin uptake inhibitors; TAU: treatment as usual; TCAs: tricyclic antidepressants.

any reason (more severe depression), so results from bias-adjusted models were considered (Appendix 6).

For less severe depression, most risk-of-bias domains were rated as low or unclear except for blinding of participants and personnel due to lack of blinding in trials assessing psychological interventions alone or in combination. For more severe depression, the majority of risk-of-bias domains were rated as low or unclear, with the exception of selective reporting bias and other bias (which included potential conflict of interest based

on the funding source). Differences in types of bias and quality ratings between less and more severe depression are attributed to the considerably larger proportion of pharmacological trials for more severe depression compared with less severe depression.

Following quality assessment using the narrative summary of the standard GRADE approach domains, evidence was judged to be of low and low-to-moderate quality for less and more severe depression, respectively (Appendix 6).

Treatment outcomes

Bias-adjusted results on the SMD of depressive symptom change of each class versus the reference treatment are shown as forest-like plots in Figs. 3 and 4, for less and more severe depression, respectively. Full results of treatment effects versus the reference treatment for all outcomes are shown in Appendix 6 (classes) and Appendix 8 (interventions). Relative effects between active treatments (classes and interventions) for all outcomes, including inconsistency model comparisons, are reported in Appendix 7.

For less severe depression, group CT/CBT was the only class with clear evidence of efficacy versus TAU on depressive symptom change (SMD -1.01, 95% CrI -1.76; -0.06). At the intervention level, group interventions (behavioural activation, problem solving, MBCT, mindfulness meditation), individual third-wave

cognitive therapy, and, marginally (with upper 95%CrI just crossing the no effect line), individual CBT (≥ 15 sessions), cognitive bibliotherapy with or without support, and a number of computerised interventions with or without support, were effective versus TAU. Most other classes and interventions showed improvements in outcomes, but 95%CrI were wide and crossed the no effect line. Group treatments showed higher effects compared with TAU regarding response and remission. No differential effects between classes were found. SSRIs and TCAs showed very small effects with high uncertainty compared with pill placebo. At the intervention level, group interventions (behavioural activation, CBT and problem solving) showed higher effect than self-help interventions with or without support.

For more severe depression, classes with evidence of efficacy on depressive symptom change versus pill

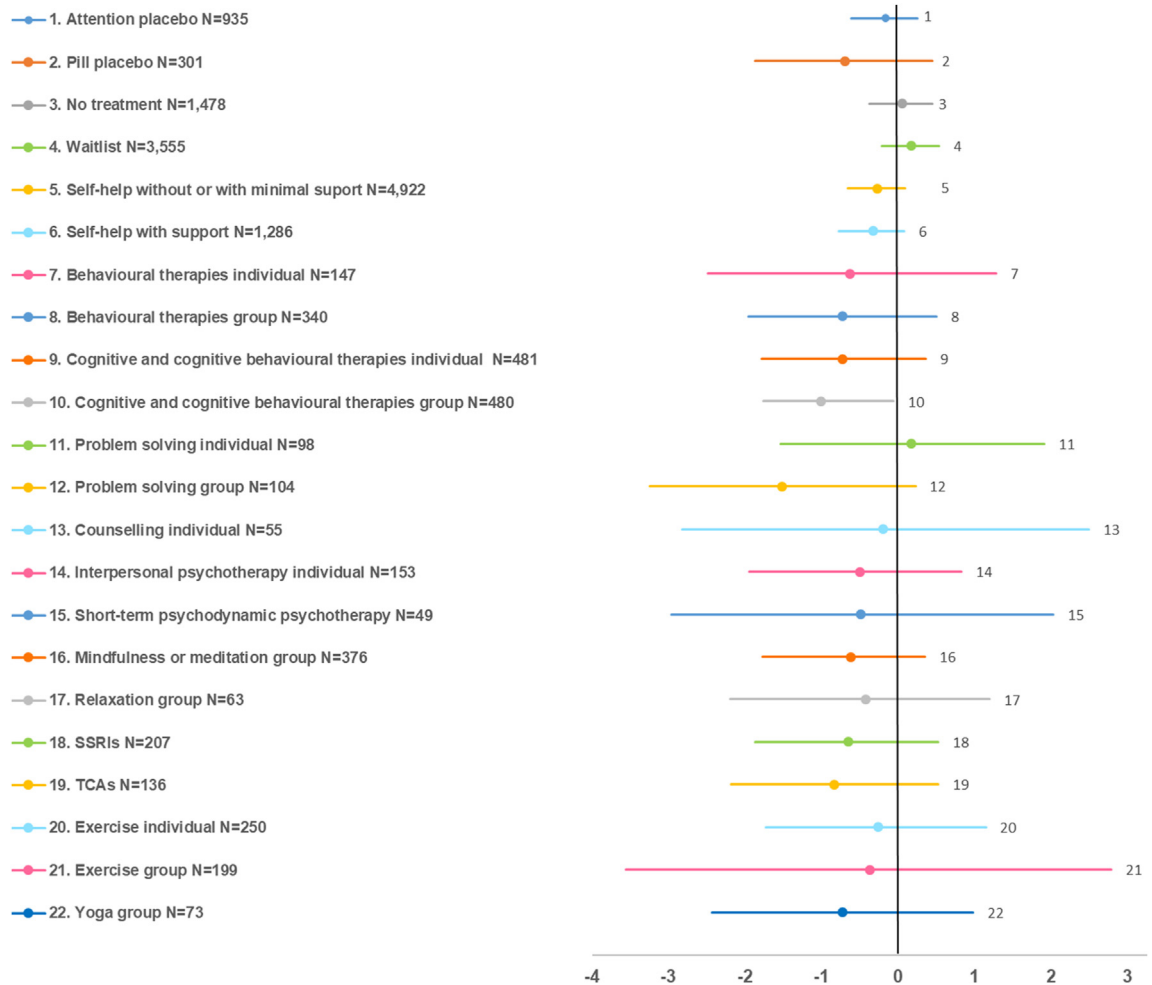


Fig. 3: Bias-adjusted forest plots of standardised mean difference (SMD) of depressive symptom change in adults with a new episode of less severe depression: effects of treatment classes versus treatment as usual (TAU, N = 815). Values on the left side of the vertical axis indicate better effect compared with TAU. Effects are shown only for treatment classes with N ≥ 50 , plus short-term psychodynamic psychotherapy (N = 49). SSRIs: selective serotonin uptake inhibitors; TCAs: tricyclic antidepressants.

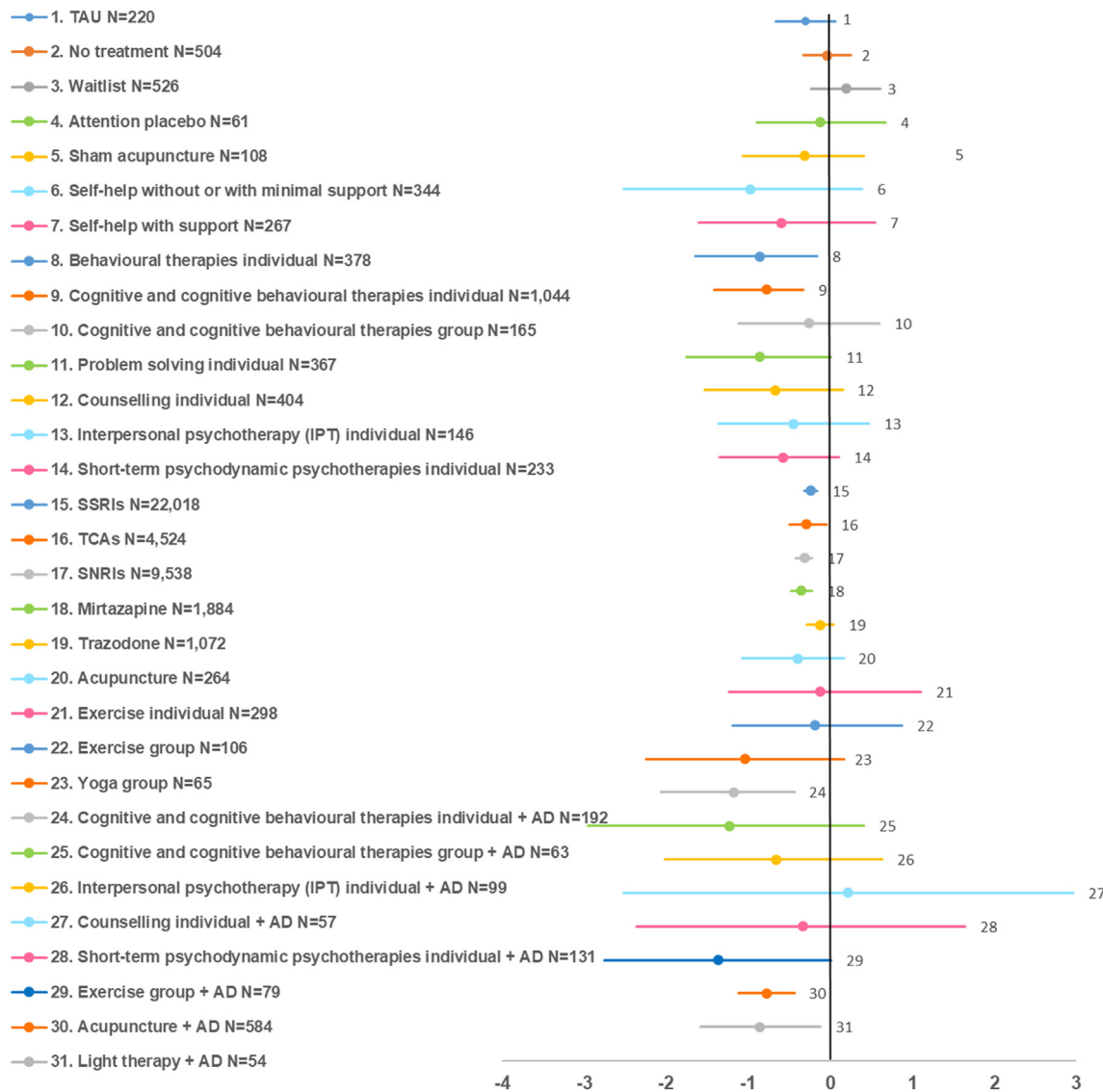


Fig. 4: Bias-adjusted forest plots of standardised mean difference (SMD) of depressive symptom change in adults with a new episode of more severe depression: effects of treatment classes versus pill placebo (N = 12,554). Values on the left side of the vertical axis indicate better effect compared with pill placebo. Effects are shown only for treatment classes with N ≥ 50. AD: antidepressant; SNRIs: serotonin and norepinephrine reuptake inhibitors; SSRIs: selective serotonin uptake inhibitors; TAU: treatment as usual; TCAs: tricyclic antidepressants.

placebo included combined individual CT/CBT with antidepressants (SMD -1.18, 95%CrI -2.07; -0.44), individual behavioural therapies (-0.86, 95%CrI -1.65; -0.16), combined light therapy with antidepressants (-0.86, 95% CrI -1.59; -0.12), combined acupuncture with antidepressants (-0.78, 95%CrI -1.12; -0.44), individual CT/CBT (-0.78, 95%CrI -1.42; -0.33), mirtazapine (-0.35, 95%CrI -0.48; -0.22), SNRIs (-0.32, 95%CrI -0.43; -0.22), TCAs (-0.29, 95%CrI -0.50; -0.05) and SSRIs (-0.24, 95% CrI -0.32; -0.16). Effective interventions belonging to other classes included cognitive bibliotherapy, computerised CBT with or without support, various individual

psychotherapies (problem solving, non-directive counselling, short-term psychodynamic psychotherapy [PDPT]), various combinations of psychological or physical interventions (group CBT, IPT, group exercise) with antidepressants, and group yoga. The majority of other classes and interventions showed improvements in outcomes, but with wide 95%CrI that crossed the no effect line. Similar results were found for response and remission. Long term PDPT alone or combined with antidepressants showed higher remission than pill placebo. Combined treatment classes (individual CT/CBT or acupuncture) with antidepressants showed higher effects than antidepressant

classes alone. At the intervention level, self-help (cognitive bibliotherapy, computerised CBT with or without support) and individual psychological interventions (problem solving, behavioural activation, CBT, non-directive counselling) were overall more effective than antidepressants. Combined traditional or electro-acupuncture with SSRIs was more effective than traditional acupuncture or antidepressants alone.

No class showed higher risk of discontinuation versus reference in either depressive symptom level. Other than SSRIs in less severe depression, all assessed antidepressant classes showed higher risk of discontinuation due to side effects compared with pill placebo.

The sensitivity analysis conducted to explore the transitivity assumption between pharmacological and non-pharmacological trial participants suggested only small changes in non-pharmacological treatment effects on the SMD after exclusion of pharmacological RCTs for less severe depression, probably because there were few pharmacological trials in this population. For more severe depression, exclusion of pharmacological trials resulted in some changes and higher uncertainty in effects and rankings of non-pharmacological treatments, apparently because most evidence in this population came from pharmacological trials (Appendix 6). Overall, non-pharmacological treatment effects were not substantially affected after exclusion of pharmacological trials.

Results from the 2023 review update

The search update identified 26,024 records. Following machine learning methods that eliminated records with high likelihood of being irrelevant, 5656 publications remained as potentially eligible, of which 165 were assessed at full text for eligibility. The selection process resulted in 16 and 10 new RCTs for less and more severe depression, respectively (see flow diagram in Appendix 9). Conclusions on the SMD (primary outcome) remained largely unchanged. However, in contrast to the original analysis for less severe depression, group yoga and self-help without support showed evidence of effect versus TAU. Group yoga also showed evidence of effect versus self-help with support in this population. At the intervention level, in addition to group interventions, individual behavioural activation, individual CBT interventions and a wider range of self-help interventions without support showed evidence of effect versus TAU. The updated analysis for more severe depression differed from the original results in that combined group exercise with antidepressants showed evidence of efficacy versus pill placebo, whereas combined light therapy with antidepressants failed to show evidence of efficacy versus pill placebo (however, interventions within this class retained evidence of efficacy versus pill placebo). At the intervention level, combined group CBT with antidepressants appeared to be more efficacious than group

CBT alone, antidepressants alone, individual exercise interventions and acupuncture. Appendix 9 and Appendix 10 show full data and results on the depressive symptom change (SMD) outcome following the review update.

Discussion

This large NMA, which was undertaken specifically to inform recommendations on first-line treatments of a new depressive episode as part of national guidance on the treatment and management of depression on adults,²¹ compared psychological, psychosocial, pharmacological, physical and combined first-line treatments for a new depressive episode stratified by symptom severity level and synthesised evidence from 676 RCTs, 105,477 participants and 63 classes in 5 efficacy and two discontinuation outcomes.

For less severe depression, predominantly the group CT/CBT class, other group interventions (e.g. behavioural activation, MBCT), individual third-wave cognitive therapy, and, marginally, individual CBT (≥ 15 sessions), cognitive bibliotherapy with or without support, and a number of computerised interventions with or without support, appeared to be effective versus TAU. Most other classes and interventions showed improvements in outcomes compared with TAU, but results were uncertain. Antidepressants did not show evidence of effect compared with either TAU or pill placebo. Results may be explained by benefits associated with group cohesion and normalization effects, learning and support from others and functioning as co-therapists^{40,41}. The lack of a clear, distinguishable effect for most other treatments versus TAU may be attributable to the increased likelihood for non-specific effects, such as natural recovery, in less severe depression. Group CT/CBT, the only class with clear evidence of effect versus TAU, was tested on $N = 480$ on the SMD.

For more severe depression, combined treatments (antidepressants combined with individual CT/CBT, light therapy, or acupuncture) were more effective than antidepressants, psychological treatments, acupuncture, or exercise alone. Individual psychological treatment classes (behavioural therapies, CT/CBT), other psychological interventions (problem solving, non-directive counselling, short-term PDPT, cognitive bibliotherapy, computerised CBT with or without support) and antidepressants (mirtazapine, SNRIs, TCAs and SSRIs) were effective versus pill placebo. Some self-help and high-intensity psychological interventions showed higher effects than antidepressants. Most other classes and interventions showed improvements in outcomes but with considerable uncertainty. Antidepressants showed higher risk of discontinuation due to side effects relative to pill placebo. Antidepressants had by far the largest evidence base (ranging from $N = 1884$ for

mirtazapine to $N = 22,018$ for SSRIs on the SMD outcome), followed by individual CT/CBT class ($N = 1044$), combined acupuncture with antidepressants ($N = 584$), and individual behavioural therapies class ($N = 378$).

Previous NMAs concluded that (a) antidepressants were more effective than pill placebo with little differences among them^{7,8}; (b) guided and unguided self-help CBT^{9,11,13} and high-intensity psychological therapies^{10–12} were more effective than TAU, waitlist and/or pill placebo; (c) individual, group, telephone, and guided self-help CBT had similar effectiveness⁹ and (d) combined psychotherapy with pharmacotherapy was more effective than psychotherapy or pharmacotherapy alone, with no differences being found between psychotherapy and pharmacotherapy.^{15,16,18} However, no previous NMA made comparisons between specific psychological and pharmacological treatment classes or interventions. Our results differ from previous NMA findings in that, according to our analyses, (a) antidepressants did not show evidence of an effect against pill placebo in less severe depression; (b) group CBT appears to be the most effective treatment for less severe depression; (c) the mode of delivery appears to impact on CBT effectiveness; and (d) some psychological interventions may be more effective than antidepressants in more severe depression (see [Appendix 7](#) for full results on comparisons between interventions). Differences between our NMA and previous NMA results may be attributable to different inclusion criteria (e.g., we included only RCTs of people treated for a new depressive episode not currently receiving treatment) and our stratification by depressive symptom severity. We also synthesised a much larger evidence base than previously published NMAs by including a wider range of treatment modalities. Compared with other published NMAs in the field, we considered a substantially larger number of distinct, clearly defined pharmacological and psychological interventions grouped into homogeneous treatment classes by taking also into account their mode of delivery (we included approximately 150 active interventions grouped into 51 active treatment classes), and, for the first time, we employed class models for analyses, which allowed us to capture subtle differences in efficacy between treatment classes and interventions within and across classes. Furthermore, we conducted analyses adjusted for bias due to small study size, which, to our knowledge, has not been done by any other NMA in the area so far.

Our conclusions on the antidepressant treatment effects in less severe depression differ from those of a recent large individual participant data (IPD) meta-analysis that assessed the effects of SSRIs for non-severe and severe depression.⁴² These differences may be further attributable to different cut-off points used to define less severe depression in our study (HAMD<16) compared with the cut-offs for the non-severe

depression in the IPD meta-analysis (HAMD<19),⁴² and the use of IPD by the other study, which enabled assessment of antidepressant treatment effects on the core symptoms of depression (expressed in the 6-item Bech HAMD sub-scale).⁴² On the other hand, our findings on antidepressant effects versus pill placebo are in line with those of an older IPD meta-analysis, which concluded that the magnitude of antidepressant effects compared with pill placebo increases with depressive symptom severity and may be overall minimal or non-existent in people with mild or moderate symptoms (defined by a HAMD score of 8–13 and 14–18, respectively).⁴³

We used strict criteria to identify a homogeneous population experiencing a new depressive episode and excluded studies where >20% of participants were already receiving treatment (most commonly antidepressants) at trial initiation or had chronic depression or depression not responding to previous treatment, and studies specifically recruiting participants with depression and physical health problems. To stratify analyses by symptom severity level we categorised studies using the study sample's mean baseline depressive symptom score. However, some study categorisations may have been incorrect if baseline score distributions were largely skewed. Moreover, the approach we used to categorise the study population into the levels of less and more severe depression is novel, using partly clinical judgement, and has not been validated, but recognises some of the inherent challenges in depression symptom scale measurements, such as (a) the default thresholds of depression symptom scales may not necessarily correspond to each other, as suggested by the use of cross-walk tables; (b) some scales may have more than one threshold sets reported in the published literature and used in routine practice; and (c) for some scales, default thresholds may overstate depressive symptom severity. The choice of threshold may have a crucial impact on the analysis results, as it affects the number of studies and trial participants as well as the availability of treatment classes and interventions in each network, and may therefore also have an impact on the estimation of bias in bias-adjusted models and the estimation of within-class variance in class models (which depends on the availability of treatment classes and interventions in the network). An alternative approach to explore the impact of depressive symptom severity on treatments effects would be by employing IPD NMA⁴⁴; however, we did not have access to IPD data.

We employed class models³⁴ to gain precision on treatment effects, connect networks disconnected at the intervention level, and aide interpretability. We used TAU (instead of the well-defined pill placebo) as the reference treatment for less severe depression, due to the very limited evidence and connectivity of pill placebo in the respective evidence networks. We reduced the

heterogeneity of the TAU node by recoding comparisons of 'intervention plus TAU versus TAU alone' as 'intervention versus no treatment', and coding TAU arms where TAU was described as a well-defined intervention or control condition accordingly. Nevertheless, the remaining TAU arms forming the TAU node may still comprise a heterogeneous condition, which is acknowledged as a limitation of the analyses for less severe depression.

Analyses for more severe depression showed some inconsistency between direct and indirect evidence. Notably, across analyses, between-study heterogeneity was lower in more severe depression compared with less severe depression. This could be explained by the larger number of pharmacological trials included in more severe depression, which typically recruit more homogeneous populations. However, exclusion of pharmacological trials had only a small impact on the SMD results, suggesting that the transitivity assumption between non-pharmacological and pharmacological trial populations is acceptable. Pairwise sub-group analysis assessing age and setting as potential effect modifiers did not find evidence suggestive of a differential effect between older (aged ≥ 60 years) versus younger (aged < 60 years) adults and between inpatient and outpatient care settings (although the latter was based on a very limited evidence base). Similar planned analyses to assess gender and ethnicity as potential effect modifiers were not possible to undertake due to lack of relevant data. In addition to those factors, it is possible that other, unknown effect modifiers (which were therefore not possible to control for) were not evenly spread across comparisons, thus increasing heterogeneity and compromising transitivity. Bias-adjusted analyses suggested the presence of small study effects favouring active versus control treatments in some outcomes, however it is unclear whether this is due to publication bias or study design aspects that may differ between smaller and larger studies. Small study effects were assumed to be common across comparisons of active interventions versus control (or non-directive counselling) in each analysis, though in reality there may be groups of active treatments (e.g. pharmacological, psychological) in which this bias is more pronounced.

In addition to the above analyses, we also conducted post-hoc pairwise meta-analyses of the available evidence, to compare their effect sizes to those obtained from the NMA. We considered the effects between the NMA and the pairwise meta-analysis to be 'significantly different' if the NMA effect estimate was not within the 95% confidence interval (CI) of the pairwise meta-analysis effect estimate. For less severe depression, out of 93 available comparisons between NMA and pairwise meta-analysis results, we found 11 comparisons (12%) where results were 'significantly different'. For more severe depression, out of 160 available comparisons between NMA and pairwise meta-analysis results, we

found 17 comparisons (11%) where effects were 'significantly different'. For most differences identified, the difference was in the magnitude rather than the direction of effect and could probably be accounted for by the smaller evidence base contributing to the pairwise effect estimates.

The quality of the NMA evidence, narratively assessed using the GRADE approach domains,³⁷ was judged to be low and low-to-moderate for less and more severe depression, respectively. This overall low quality of the evidence in this research area has been reported in other published NMAs.^{7,10,18} Originally, a threshold analysis was planned to test the robustness of the NMA-based treatment recommendations to potential biases or sampling variation in included evidence.⁴⁵ However, this analysis was ultimately not considered feasible or helpful to conduct due to the lack of clear decision-making rules linking recommendations to the NMA results, as other pragmatic criteria were factored in (discussed below). CINeMA, a method used to evaluate the confidence in the NMA results⁴⁶ was also not possible to undertake due to the implementation of class models.

Our systematic review included evidence published in English language unless data could be extracted from an existing review.⁷ Evidence suggests that use of language restrictions in systematic review-based meta-analyses in conventional medicine does not introduce systematic bias.⁴⁷ Final NICE guideline evidence searches were conducted in June 2020, due to the complexity and time required to run the analyses prior to guideline publication. Given the possibility of new evidence becoming available since then, we updated our searches and analyses on the depressive symptom change (SMD) outcome in November 2023, using novel machine learning methods to speed up the study selection process. Results remained largely unchanged. In less severe depression, group yoga, self-help without support and some individual psychological interventions (behavioural activation, CBT) emerged as efficacious versus TAU. Results for group yoga were attributable to the inclusion of one small RCT (N = 62) with a substantial positive effect for laughter group yoga (Armat 2022). In more severe depression, combined group exercise with antidepressants emerged as efficacious versus pill placebo (albeit based on a more limited evidence base than other efficacious classes in this population), whereas light therapy combined with antidepressants failed to show efficacy as in the original analysis, which is likely attributable to its limited evidence base, which apparently could not ensure stability in the results. The similarity between the updated and original results on the SMD outcome is not surprising given the particularly large evidence base that was synthesised in our original analysis (so that a small body of newly identified evidence was unlikely to shift conclusions to a considerable extent) and increases our confidence in our original conclusions. A search update

conducted in July 2024 identified 13 eligible studies for our review and NMA of the primary outcome of the SMD of depressive symptom change (Appendix 11). The newly identified evidence for less severe depression (9 RCTs) might potentially reduce the uncertainty in findings for group exercise, group mindfulness or meditation, self-help with support, and IPT, possibly weaken the findings for group CBT, and strengthen the updated findings for yoga and self-help, although the size of this new evidence is very small compared with the evidence base synthesised in our NMA and thus unlikely to be influential to our reported results. Some new evidence was also found for light therapy in less severe depression. For more severe depression, there was very limited evidence (4 RCTs) on self-help (with or without support) and SSRIs that is highly unlikely to alter our reported conclusions. However, it is likely that, as new evidence accumulates, especially for currently less well-researched treatments, some treatments with currently uncertain findings may show evidence of efficacy in the future.

In terms of developing clinical recommendations, the findings presented here were not the only data considered by the guideline committee. The full guideline process is described elsewhere²¹ but, in making recommendations, the guideline committee took into account the size, uncertainty and limitations of both the clinical and cost-effectiveness evidence, results of pairwise meta-analysis of outcomes at follow-up, quality of life and functioning data, qualitative evidence on the facilitators and barriers to treatment choice, side effects and withdrawal symptoms associated with antidepressants, the applicability of the evidence to the UK context (e.g. evidence on acupuncture came predominantly from Chinese studies and there may be differences between Chinese and Western-style acupuncture techniques), implementation issues with regard to current structure of psychological treatment services in England, and the need to provide a wide range of interventions to meet individual needs and support patient choice.

Based on the NMA findings and the above considerations, the NICE updated guideline on Depression in adults²¹ recommends a wide range of first-line treatments for adults with a new episode of depression, arranged in a suggested order in which they should be considered within a shared decision-making context (see NICE visual summaries for new episodes of [less severe depression](#) and [more severe depression](#)). Recommendations generally reflect current practice, but may reduce variation in practice across the National Health Service in England. Further research was recommended on peer support, combined acupuncture with antidepressants, differential effects of psychological interventions and withdrawal effects of antidepressants for which evidence was more limited, less applicable, or uncertain. Finally, this analysis is limited through the

essentially unitary concept of ‘depression’ assessed within clinical trials, despite clinical wisdom and research evidence continuing to point out its heterogeneity.^{48–50} Clearer subtyping, associated with causal mechanisms (from basic biomarkers to higher order reinforcement learning), may provide enhanced predictive signals for more targeted clinical interventions and we recommend this as an area for future research investment.

Contributors

OMV, IM, SP, SD, NJW and HP drafted the study protocol. SS and SA performed the literature searches. OMV and KJMOD assessed the eligibility of the studies for inclusion, extracted the data and assessed the risk of bias. SD, NJW, HP, CHD, EK, and DMC developed the statistical code and ran the statistical analyses. JT performed the machine learning methods of the 2023 review update and 2024 search update. All authors (IM, OMV, HP, NJW, SD, EW, NN, CHD, EK, HE, DMC, KJMOD, SS, SA, JT, NK, SP) contributed to the design of the study and the selection of studies for inclusion in the analysis and contributed to the interpretation of the findings. IM drafted the first draft of the manuscript, which all authors critically reviewed and contributed to, and approved its final version submitted for publication. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Data sharing statement

All data used in the study are within the paper and its Supporting Information files.

Declaration of interests

HE, KJMOD, IM, OMV, SA, SS and SP received support from NICE for the submitted work. CHD, DMC, HP, NJW, EK, and SD received support from the NICE Guidelines TSU for the submitted work. EW, NK and NN declared the following interests based on NICE’s policy on conflicts of interests: <https://www.nice.org.uk/guidance/ng222/documents/register-of-interests>. The authors report no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

This work was funded by the National Institute for Health and Care Excellence (NICE). IM, OMV and SP are employees of University College London which has received funding from NICE to develop clinical guidelines, including the submitted work. HE, KJMOD and SA are employees of NICE. SS was employed by the National Guideline Alliance (NGA) at the Royal College of Obstetricians and Gynaecologists (RCOG), which received funding from NICE for the development of clinical, public health and social care guidelines, including the submitted work. CHD, DMC, EK, HP, NJW and SD were supported by the NICE Guidelines Technical Support Unit (TSU), University of Bristol, which is funded by the NICE Centre for Guidelines to support the development of clinical guidelines, including the submitted work. NK is an employee of the University of Manchester which was not commissioned or funded as part of the guideline development process. EW is an employee of the University of Exeter, which was not commissioned or funded as part of the guideline development process. JT is an employee of the EPPI-Centre, Social Research Institute at University College London, which was not commissioned or funded for this work. NN is an employee of the University of Nottingham, which was not commissioned or funded as part of the guideline development process. The views expressed in this publication are those of the authors and not necessarily those of UCL, RCOG, University of Bristol or NICE. The funder of the study had no further role in the study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

National Institute for Health and Care Excellence (2022) Depression in adults: Treatment and management. Available from: <https://www.nice.org.uk/guidance/ng222>.

We thank other members of the guideline committee for the NICE guideline on 'Depression in adults: treatment and management' (2022 update) for their contributions to this work. The committee included clinical academics, healthcare professionals, service commissioners, and lay people with expertise and experience in the field of depression. The full list of committee members can be found [here](#). We thank Tony Ades at University of Bristol for invaluable advice during study development, and Kate Ren at the University of Sheffield for her useful comments and detailed quality assurance of this work. We also thank colleagues at University College London and the Royal College of Obstetricians and Gynaecologists that provided technical input and administrative support at various stages of the guideline development.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2024.102780>.

References

- Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2163-2196.
- McCrone P, Dhanasiri S, Patel A, Knapp M, Lawton-Smith S. *Paying the price. The cost of mental health care in England to 2026*. London: King's Fund; 2008.
- Greenberg PE, Fournier AA, Sisitsky T, et al. The economic burden of adults with major depressive disorder in the United States (2010 and 2018). *Pharmacoeconomics*. 2021;39(6):653-665.
- Cuijpers P, Dekker J, Hollon SD, Andersson G. Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: a meta-analysis. *J Clin Psychiatry*. 2009;70(9):1219-1229.
- Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol*. 2008;76(6):909-922.
- Cuijpers P, van Straten A, van Oppen P, Andersson G. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J Clin Psychiatry*. 2008;69(11):1675-1685. quiz 839-41.
- 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(10128):1357-1366.
- Linde K, Kriston L, Rucker G, et al. Efficacy and acceptability of pharmacological treatments for depressive disorders in primary care: systematic review and network meta-analysis. *Ann Fam Med*. 2015;13(1):69-79.
- Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatr*. 2019;76(7):700-707.
- 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(2):283-293.
- Linde K, Rucker G, Sigterman K, et al. Comparative effectiveness of psychological treatments for depressive disorders in primary care: network meta-analysis. *BMC Fam Pract*. 2015;16:103.
- 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(5):e1001454.
- Karyotaki E, Efthimiou O, Miguel C, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatr*. 2021;78(4):361-371.
- Ciharova M, Furukawa TA, Efthimiou O, et al. Cognitive restructuring, behavioral activation and cognitive-behavioral therapy in the treatment of adult depression: a network meta-analysis. *J Consult Clin Psychol*. 2021;89(6):563-574.
- 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(1):92-107.
- Cuijpers P, Oud M, Karyotaki E, et al. Psychologic treatment of depression compared with pharmacotherapy and combined treatment in primary care: a network meta-analysis. *Ann Fam Med*. 2021;19(3):262-270.
- Jiang X, Luo Y, Chen Y, et al. Comparative efficacy of multiple therapies for the treatment of patients with subthreshold depression: a systematic review and network meta-analysis. *Front Behav Neurosci*. 2021;15:755547.
- Gartlehner G, Dobrescu A, Chapman A, et al. Nonpharmacologic and pharmacologic treatments of adult patients with major depressive disorder: a systematic review and network meta-analysis for a clinical guideline by the American College of Physicians. *Ann Intern Med*. 2023;176(2):196-211.
- Noetel M, Sanders T, Gallardo-Gomez D, et al. Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials. *BMJ*. 2024;384:e075847.
- Zimmerman M, Martinez JH, Friedman M, Boerescu DA, Attiullah N, Toba C. How can we use depression severity to guide treatment selection when measures of depression categorize patients differently? *J Clin Psychiatry*. 2012;73(10):1287-1291.
- National Institute for Health and Care Excellence. *Depression in adults: treatment and management*; 2022. Available at: <https://www.nice.org.uk/guidance/ng222>.
- 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(11):777-784.
- Carmody TJ, Rush AJ, Bernstein I, et al. The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *Eur Neuropsychopharmacol*. 2006;16(8):601-611.
- Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573-583.
- Uher R, Farmer A, Maier W, et al. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med*. 2008;38(2):289-300.
- Wahl I, Lowe B, Bjorner JB, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol*. 2014;67(1):73-86.
- Manea L, Gilbody S, McMillan D. Optimal cut-off score for diagnosing depression with the Patient Health Questionnaire (PHQ-9): a meta-analysis. *CMAJ*. 2012;184(3):E191-E196.
- Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord*. 2004;78(2):131-140.
- Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J Gen Pract*. 2008;58(546):32-36.
- Higgins JPT, Thomas J, Chandler J, et al. *Cochrane handbook for systematic reviews of interventions*. 2nd ed. Chichester (UK): John Wiley & Sons; 2019.
- Spiegelhalter D, Thomas A, Best N, Lunn DJ. *WinBUGS user manual: version 1.4*. Cambridge: MRC Biostatistics Unit; 2003.
- Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. *The BUGS book*. Boca Raton, FL: CRC Press; 2013.
- Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-A Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput*. 2000;10:325-337.
- Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network meta-analysis for decision-making*. Hoboken, NJ: Wiley; 2018.
- Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. *Evid Based Ment Health*. 2015;18(2):40-46.
- Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making*. 2013;33(5):641-656.
- Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924-926.
- Dias S, Welton NJ, Marinho VCC, Salanti G, Higgins JPT, Ades AE. Estimation and adjustment of bias in randomised evidence by

- using Mixed Treatment Comparison Meta-analysis. *J R Stat Soc Ser A Stat Soc.* 2010;173(3):613–629.
- 39 Moreno SG, Sutton AJ, Turner EH, et al. Novel methods to deal with publication biases: secondary analysis of antidepressant trials in the FDA trial registry database and related journal publications. *BMJ.* 2009;339:b2981.
- 40 Vickery A. 'It's made me feel less isolated because there are other people who are experiencing the same or very similar to you': men's experiences of using mental health support groups. *Health Soc Care Community.* 2022;30(6):2383–2391.
- 41 Thimm JC, Antonsen L. Effectiveness of cognitive behavioral group therapy for depression in routine practice. *BMC Psychiatry.* 2014;14:292.
- 42 Hieronymus F, Lisinski A, Nilsson S, Eriksson E. Influence of baseline severity on the effects of SSRIs in depression: an item-based, patient-level post-hoc analysis. *Lancet Psychiatry.* 2019;6(9):745–752.
- 43 Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA.* 2010;303(1):47–53.
- 44 Riley RD, Dias S, Donegan S, et al. Using individual participant data to improve network meta-analysis projects. *BMJ Evid Based Med.* 2023;28(3):197–203.
- 45 Phillippo DM, Dias S, Welton NJ, Caldwell DM, Taske N, Ades AE. Threshold analysis as an alternative to GRADE for assessing confidence in guideline recommendations based on network meta-analyses. *Ann Intern Med.* 2019;170(8):538–546.
- 46 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(4):e1003082.
- 47 Morrison A, Polisena J, Huserau D, et al. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *Int J Technol Assess Health Care.* 2012;28(2):138–144.
- 48 Alshehri T, Mook-Kanamori DO, Willems van Dijk K, et al. Metabolomics dissection of depression heterogeneity and related cardiometabolic risk. *Psychol Med.* 2023;53(1):248–257.
- 49 Nguyen TD, Harder A, Xiong Y, et al. Genetic heterogeneity and subtypes of major depression. *Mol Psychiatry.* 2022;27(3):1667–1675.
- 50 Buch AM, Liston C. Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. *Neuropsychopharmacology.* 2021;46(1):156–175.