

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



Andrea Cipriani^{*}, Toshi A Furukawa^{*}, Georgia Salanti^{*}, Anna Chaimani, Lauren Z Atkinson, Yusuke Ogawa, Stefan Leucht, Henricus G Ruhe, Erick H Turner, Julian P T Higgins, Matthias Egger, Nozomi Takeshima, Yu Hayasaka, Hissei Imai, Kiyomi Shinohara, Aran Tajika, John P A Ioannidis, John R Geddes



Summary

Background Major depressive disorder is one of the most common, burdensome, and costly psychiatric disorders worldwide in adults. Pharmacological and non-pharmacological treatments are available; however, because of inadequate resources, antidepressants are used more frequently than psychological interventions. Prescription of these agents should be informed by the best available evidence. Therefore, we aimed to update and expand our previous work to compare and rank antidepressants for the acute treatment of adults with unipolar major depressive disorder.

Methods We did a systematic review and network meta-analysis. We searched Cochrane Central Register of Controlled Trials, CINAHL, Embase, LILACS database, MEDLINE, MEDLINE In-Process, PsycINFO, the websites of regulatory agencies, and international registers for published and unpublished, double-blind, randomised controlled trials from their inception to Jan 8, 2016. We included placebo-controlled and head-to-head trials of 21 antidepressants used for the acute treatment of adults (≥ 18 years old and of both sexes) with major depressive disorder diagnosed according to standard operationalised criteria. We excluded quasi-randomised trials and trials that were incomplete or included 20% or more of participants with bipolar disorder, psychotic depression, or treatment-resistant depression; or patients with a serious concomitant medical illness. We extracted data following a predefined hierarchy. In network meta-analysis, we used group-level data. We assessed the studies' risk of bias in accordance to the Cochrane Handbook for Systematic Reviews of Interventions, and certainty of evidence using the Grading of Recommendations Assessment, Development and Evaluation framework. Primary outcomes were efficacy (response rate) and acceptability (treatment discontinuations due to any cause). We estimated summary odds ratios (ORs) using pairwise and network meta-analysis with random effects. This study is registered with PROSPERO, number CRD42012002291.

Findings We identified 28 552 citations and of these included 522 trials comprising 116 477 participants. In terms of efficacy, all antidepressants were more effective than placebo, with ORs ranging between 2.13 (95% credible interval [CrI] 1.89–2.41) for amitriptyline and 1.37 (1.16–1.63) for reboxetine. For acceptability, only agomelatine (OR 0.84, 95% CrI 0.72–0.97) and fluoxetine (0.88, 0.80–0.96) were associated with fewer dropouts than placebo, whereas clomipramine was worse than placebo (1.30, 1.01–1.68). When all trials were considered, differences in ORs between antidepressants ranged from 1.15 to 1.55 for efficacy and from 0.64 to 0.83 for acceptability, with wide CrIs on most of the comparative analyses. In head-to-head studies, agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, venlafaxine, and vortioxetine were more effective than other antidepressants (range of ORs 1.19–1.96), whereas fluoxetine, fluvoxamine, reboxetine, and trazodone were the least efficacious drugs (0.51–0.84). For acceptability, agomelatine, citalopram, escitalopram, fluoxetine, sertraline, and vortioxetine were more tolerable than other antidepressants (range of ORs 0.43–0.77), whereas amitriptyline, clomipramine, duloxetine, fluvoxamine, reboxetine, trazodone, and venlafaxine had the highest dropout rates (1.30–2.32). 46 (9%) of 522 trials were rated as high risk of bias, 380 (73%) trials as moderate, and 96 (18%) as low; and the certainty of evidence was moderate to very low.

Interpretation All antidepressants were more efficacious than placebo in adults with major depressive disorder. Smaller differences between active drugs were found when placebo-controlled trials were included in the analysis, whereas there was more variability in efficacy and acceptability in head-to-head trials. These results should serve evidence-based practice and inform patients, physicians, guideline developers, and policy makers on the relative merits of the different antidepressants.

Funding National Institute for Health Research Oxford Health Biomedical Research Centre and the Japan Society for the Promotion of Science.

Copyright © The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

Published Online
February 21, 2018
[http://dx.doi.org/10.1016/S0140-6736\(17\)32802-7](http://dx.doi.org/10.1016/S0140-6736(17)32802-7)

See Online/Comment
[http://dx.doi.org/10.1016/S0140-6736\(18\)30421-5](http://dx.doi.org/10.1016/S0140-6736(18)30421-5)

^{*}Joint first authors

Department of Psychiatry, University of Oxford, Oxford, UK (A Cipriani MD, L Z Atkinson MSc, H G Ruhe PhD, Prof J R Geddes MD); Oxford Health NHS Foundation Trust (A Cipriani, Prof J R Geddes) and Oxford Centre for Human Brain Activity, Department of Psychiatry (L Z Atkinson), Warneford Hospital, Oxford, UK; Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine and School of Public Health, Kyoto, Japan

(Prof T A Furukawa MD, Y Ogawa MD, N Takeshima MD, Y Hayasaka MD, H Imai MD, K Shinohara MD, A Tajika MD); Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland (G Salanti PhD, Prof M Egger PhD); School of Medicine, Paris Descartes University, Paris, France (A Chaimani PhD); INSERM, UMR1153 Epidemiology and Statistics, Sorbonne Paris Cité Research Center, METHODS Team, Paris, France (A Chaimani); Cochrane France, Paris, France (A Chaimani); Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany (Prof S Leucht MD); Department of Psychiatry, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands (H G Ruhe); Behavioral Health and Neurosciences Division, VA Portland Health Care System, Portland, OR, USA (E H Turner MD); Department of

Psychiatry and Department of Pharmacology, Oregon Health & Science University, Portland, OR, USA (E H Turner); School of Social and Community Medicine, University of Bristol, Bristol, UK (Prof J P T Higgins PhD); Department of Medicine, Department of Health Research and Policy, Department of Biomedical Data Science, and Department of Statistics, Stanford University, Stanford, CA, USA (Prof J P A Ioannidis MD); and Meta-Research Innovation Center at Stanford, Stanford University, Stanford, CA, USA (J P A Ioannidis)

Correspondence to: Dr Andrea Cipriani, Department of Psychiatry, University of Oxford, Oxford, OX3 7JX, UK andrea.cipriani@psych.ox.ac.uk

Research in context

Evidence before this study

Antidepressants are routinely used worldwide for the treatment of major depressive disorder, which is one of the most important global health challenges; however, in the scientific literature, there remains considerable debate about both their effectiveness as a group, and the potential differences in effectiveness and tolerability between individual drugs. With the marketing of new antidepressants and increasing numbers of trials published every year, an updated systematic review and network meta-analysis was required to synthesise the evidence in this important clinical area.

Added value of this study

This network meta-analysis represents a major update and extension of our previous study, which addressed 12 antidepressants with data for head-to-head comparisons only, and provides the best currently available evidence base to guide the choice about pharmacological treatment for adults with acute

major depressive disorder. We now include a more comprehensive list of 21 antidepressants and placebo, consider three new clinical outcome measures and many potential effect modifiers, and use the most advanced statistical methodology for network meta-analysis to date.

Implications of all the available evidence

Our findings should inform clinical guidelines and assist the shared decision making process between patients, carers, and clinicians in routine practice on selecting the most appropriate antidepressant for adults with acute major depressive disorder. Future research should seek to extend network meta-analysis to combine aggregate and individual-patient data from trials in a so-called individual-patient data network meta-analysis. This analysis will allow the prediction of personalised clinical outcomes, such as early response or specific side-effects, and the estimate of comparative efficacy at multiple timepoints.

Introduction

Psychiatric disorders account for 22.8% of the global burden of diseases.¹ The leading cause of this disability is depression, which has substantially increased since 1990, largely driven by population growth and ageing.² With an estimated 350 million people affected globally, the economic burden of depressive disorders in the USA alone has been estimated to be more than US\$210 billion, with approximately 45% attributable to direct costs, 5% to suicide-related costs, and 50% to workplace costs.³ This trend poses a substantial challenge for health systems in both developed and developing countries, with the need to treat patients, optimise resources, and improve overall health care in mental health.

Grouped into various classes of drugs with slightly different mechanisms of action, antidepressants are widely used treatments for major depressive disorder, which are available worldwide. However, there is a long-lasting debate and concern about their efficacy and effectiveness, because short-term benefits are, on average, modest; and because long-term balance of benefits and harms is often understudied.⁴ Therefore, innovation in psychopharmacology is of crucial importance, but the identification of new molecular targets is difficult, primarily because of the paucity of knowledge about how antidepressants work.⁵ In routine practice, clinicians have a wide choice of individual drugs and they need good evidence to make the best choice for each individual patient. Network meta-analyses of existing datasets make it possible to estimate comparative efficacy, summarise and interpret the wider picture of the evidence base, and to understand the relative merits of the multiple interventions.⁶ Therefore, in this study, we aimed to do a systematic review and network meta-analysis to inform clinical

practice by comparing different antidepressants for the acute treatment of adults with unipolar major depressive disorder.

Methods

Search strategy and selection criteria

We did a systematic review and network meta-analysis. We searched the Cochrane Central Register of Controlled Trials, CINAHL, Embase, LILACS database, MEDLINE, MEDLINE In-Process, PsycINFO, AMED, the UK National Research Register, and PSYINDEX from the date of their inception to Jan 8, 2016, with no language restrictions. We used the search terms “depress*” OR “dysthymi*” OR “adjustment disorder*” OR “mood disorder*” OR “affective disorder” OR “affective symptoms” combined with a list of all included antidepressants.

We included double-blind, randomised controlled trials (RCTs) comparing antidepressants with placebo or another active antidepressant as oral monotherapy for the acute treatment of adults (≥ 18 years old and of both sexes) with a primary diagnosis of major depressive disorder according to standard operationalised diagnostic criteria (Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-5, and ICD-10). We considered only double-blind trials because we included placebo in the network meta-analysis, and because this study design increases methodological rigour by minimising performance and ascertainment biases.⁷ Additionally, we included all second-generation antidepressants approved by the regulatory agencies in the USA, Europe, or Japan: agomelatine, bupropion, citalopram, desvenlafaxine, duloxetine, escitalopram, fluoxetine, fluvoxamine, levomilnacipran, milnacipran, mirtazapine, paroxetine, reboxetine, sertraline, venlafaxine, vilazodone, and vortioxetine. To inform clinical practice

globally, we selected the two tricyclics (amitriptyline and clomipramine) included in the WHO Model List of Essential Medicines). We also included trazodone and nefazodone, because of their distinct effect and tolerability profiles. Additionally, we included trials that allowed rescue medications so long as they were equally provided among the randomised groups. We included data only for drugs within the therapeutic range (appendix pp 133, 134). Finally, we excluded quasi-randomised trials and trials that were incomplete or included 20% or more of participants with bipolar disorder, psychotic depression, or treatment-resistant depression; or patients with a serious concomitant medical illness.

The electronic database searches were supplemented with manual searches for published, unpublished, and ongoing RCTs in international trial registers, websites of drug approval agencies, and key scientific journals in the field.⁸ For example, we searched ClinicalTrials.gov using the search term “major depressive disorder” combined with a list of all included antidepressants. We contacted all the pharmaceutical companies marketing antidepressants and asked for supplemental unpublished information about both premarketing and post-marketing studies, with a specific focus on second-generation antidepressants. We also contacted study authors and drug manufacturers to supplement incomplete reports of the original papers or provide data for unpublished studies.

Six pairs of investigators (ACi, TAF, LZA, SL, HGR, YO, NT, YH, EHT, HI, KS, and AT) independently selected the studies, reviewed the main reports and supplementary materials, extracted the relevant information from the included trials, and assessed the risk of bias. Any discrepancies were resolved by consensus and arbitration by a panel of investigators within the review team (ACi, TAF, LZA, EHT, and JRG).

The full protocol of this network meta-analysis has been published.⁸

Outcomes

Our primary outcomes were efficacy (response rate measured by the total number of patients who had a reduction of $\geq 50\%$ of the total score on a standardised observer-rating scale for depression) and acceptability (treatment discontinuation measured by the proportion of patients who withdrew for any reason).⁸ All-cause discontinuation was used as a measure for the acceptability of treatments, because it encompasses efficacy and tolerability.⁹ Secondary outcomes were endpoint depression score, remission rate, and the proportion of patients who dropped out early because of adverse events. When depressive symptoms had been measured with more than one standardised rating scale, we used a predefined hierarchy, based on psychometric properties and consistency of use across included trials.⁸ In the absence of information or supplemental data from the authors, response rate was calculated according to

a validated imputation method.¹⁰ We recorded the outcomes as close to 8 weeks as possible for all analyses.⁹ If information at 8 weeks was not available, we used data ranging between 4 and 12 weeks (we gave preference to the timepoint closest to 8 weeks; if equidistant, we took the longer outcome). We checked trial protocols where available and compared published with unpublished data. We extracted data following a predefined hierarchy described in our protocol and gave priority to unpublished information in case of disagreement.⁸

Data analysis

For studies published more than once (ie, duplicates), we included only the report with the most informative and complete data. Full details of the applied statistical approaches are provided in the protocol.⁸ We estimated summary odds ratios (ORs) for dichotomous outcomes and standardised mean differences (SMD, Cohen's *d*) for continuous outcomes using pairwise and network meta-analysis. In network meta-analysis, we used group-level data; the binomial likelihood was used for dichotomous outcomes and the normal likelihood for continuous outcomes. The study effect sizes were then synthesised using a random-effects network meta-analysis model. We accounted for the correlations induced by multi-group studies by using multivariate distributions. The variance in the random-effects distribution (heterogeneity variance) was considered to measure the extent of across-study and within-comparison variability on treatment effects. Additionally, in network meta-analysis, we assumed that the amount of heterogeneity was the same for all treatment comparisons. To assess the amount of heterogeneity, we compared the posterior distribution of the estimated heterogeneity variance with its predictive distribution.¹¹ To rank the treatments for each outcome, we used the surface under the cumulative ranking curve (SUCRA) and the mean ranks.¹² The transitivity assumption underlying network meta-analysis was evaluated by comparing the distribution of clinical and methodological variables that could act as effect modifiers across treatment comparisons.⁸ We did a statistical evaluation of consistency (ie, the agreement between direct and indirect evidence) using the design-by-treatment test¹³ and by separating direct evidence from indirect evidence.¹⁴

We assessed the studies' risk of bias in accordance to the Cochrane Handbook for Systematic Reviews of Interventions. Additionally, we assessed the certainty of evidence contributing to network estimates of the main outcomes with the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework.¹⁵

We evaluated whether treatment effects for the two primary outcomes were robust in subgroup analyses and network meta-regression using study year, sponsorship, depressive severity at baseline, dosing schedule, study precision (ie, small study effect), and novelty effect.¹⁶ The appendix (pp 133–36) summarises

For the 20th WHO Model List of Essential Medicines see http://www.who.int/medicines/publications/essentialmedicines/20th_EML2017.pdf?ua=1

See Online for appendix

For the Cochrane Handbook for Systematic Reviews of Interventions see <http://handbook-5-1.cochrane.org/>

the definition of covariates. The sensitivity of our conclusions was evaluated by analysing the dataset with the following restrictions: studies with reported response rate, studies using accepted doses in all groups, studies with unpublished data, multi-centre studies, and head-to-head studies. We used comparison-adjusted funnel plots to investigate whether results in imprecise trials differ from those in more precise trials.¹⁷

We fitted all models in OpenBUGS (version 3.2.2)¹⁸ using the binomial likelihood for dichotomous outcomes, uninformative prior distributions for the treatment effects, and a minimally informative prior distribution for the common heterogeneity SD. We assumed uninformative priors—ie, $N(0,1000)$ —for all meta-regression coefficients. Convergence of models was ensured by visual inspection of three chains and after considering the Brooks–Gelman–Rubin diagnostic. The codes of analyses, statistical details of the meta-analysis, and meta-regression models are presented in the appendix (pp 182, 183). Statistical evaluation of inconsistency and production of network graphs and result figures were done using the network and network graphs packages in Stata (version 14.2).¹⁹ Network meta-analyses of the primary outcomes were duplicated using the netmeta 0.9-6 package in R (version 3.4.0).²⁰ The appendix (p 289) lists the changes to the original protocol. The study was done from March 12, 2012, to June 4, 2016, and data analysis was done from June 5, 2016, to Sept 18, 2017.

This study is registered with PROSPERO, number CRD42012002291.

Data sharing

With the publication of this Article, the full dataset will be freely available online in Mendeley Data, a secure online repository for research data, which allows archiving of any file type and assigns a permanent and unique digital object identifier (DOI) so that the files can be easily referenced (DOI:10.17632/83rthbp8ys.2).

Role of the funding source

The funder of this study had no role in study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit for publication. ACi, TAF, GS, ACh, LZA, and YO had full access to all the data, and ACi was responsible for the decision to submit for publication.

Results

28552 citations were identified by the search and 680 potentially eligible articles were retrieved in full text (figure 1). We included 421 trials from the database search, 86 unpublished studies from trial registries and pharmaceutical company websites, and 15 from personal communication or hand-searching other review articles. Overall, 522 double-blind, parallel, RCTs (comprising 116477 patients) done between 1979 and 2016, and comparing 21 antidepressants or placebo were included

in the analysis (appendix pp 6–64). The appendix (pp 65–114) summarises the characteristics of included studies. The mean study sample size was 224 participants (SD 186). In total, 87052 participants were randomly assigned to an active drug and 29425 were randomly assigned to placebo. The mean age was 44 years (SD 9) for both men and women; 38404 (62.3%) of 61681 of the sample population were women. The median duration of the acute treatment was 8 weeks (IQR 6–8). 243 (47%) of 522 studies randomly assigned participants to three or more groups, and 304 (58%) of 522 were placebo-controlled trials. 391 (83%) of 472 were multi-centre studies and 335 (77%) of 437 studies recruited outpatients only. 252 (48%) of 522 trials recruited patients from North America, 37 (7%) from Asia, and 140 (27%) from Europe (59 [11%] trials were cross-continental and the remaining 34 [7%] were either from other regions or did not specify). The great majority of patients had moderate-to-severe major depressive disorder, with a mean reported baseline severity score on the Hamilton Depression Rating Scale 17-item of 25.7 (SD 3.97) among 464 (89%) of 522 studies. Response rate was imputed in 20608 (17.7%) of 116447 cases. Rescue medications (typically benzodiazepines or other sedative hypnotics) were allowed in 187 (36%) of 522 studies. 409 (78%) of 522 studies were funded by pharmaceutical companies. We retrieved unpublished information for 274 (52%) of the included trials. Consistent with the study protocol, the primary analysis was based on the 474 studies (comprising 106966 patients) that used drugs within the licensed dose range (ie, the dosage approved by the regulatory agencies in the USA and Europe; appendix pp 133, 134).

Figure 2 shows the network of eligible comparisons for efficacy and acceptability. All antidepressants, except milnacipran, had at least one placebo-controlled trial. Only levomilnacipran was not directly compared with at least another active drug in any of the networks. The appendix (pp 139–44) provides detailed results of pairwise meta-analyses. Figure 3 shows the network meta-analysis' results for the primary outcomes. In terms of efficacy (432 RCTs, comprising 102443 patients), all antidepressants were more effective than placebo, with ORs ranging between 2.13 (95% credible interval [CrI] 1.89–2.41) for amitriptyline and 1.37 (1.16–1.63) for reboxetine. In terms of acceptability (422 RCTs, comprising 99787 patients), agomelatine (OR 0.84, 95% CrI 0.72–0.97) and fluoxetine (0.88, 0.80–0.96) were associated with fewer dropouts than placebo; by contrast, clomipramine was worse than placebo (1.30, 1.01–1.68).

The relative efficacy of antidepressants compared with placebo is also shown for remission (appendix pp 152, 153). The random-effects summary SMD for all antidepressants was 0.30 (95% CrI 0.26–0.34; $p < 0.0001$; appendix pp 150, 151). In terms of dropouts due to adverse events, all active drugs were associated with higher withdrawal rates than placebo with ORs ranging

between 1.64 and 4.44, and 95% CrI excluding the null, except agomelatine (OR 1.21, 95% CrI 0.94–1.56; appendix pp 154–55). For the full results of the secondary outcomes see the appendix (pp 150–55).

In the analysis of response rate, 8% of the loops were inconsistent (17 of 219 loops; p value of the design by treatment test was 0.063), and also 8% of the loops were inconsistent for dropouts (16 of 210 loops; $p=0.219$). The median heterogeneity variances were estimated at 0.044 (95% CrI 0.028–0.063) for response and 0.040 (0.023–0.062) for dropout, suggesting moderate-to-low heterogeneity. Subgroup meta-regression analyses revealed that the use of placebo in trials was the strongest explanation of heterogeneity and inconsistency in those evaluated. Exclusion of placebo-controlled trials resulted

in a 24% relative reduction in heterogeneity variance for response and 45% for dropout. Additionally, we found that smaller and older studies presented larger effects of the active interventions versus placebo (in particular for amitriptyline, bupropion, fluoxetine, and reboxetine; appendix pp 182–96). The year of randomisation or study precision did not materially impact on the relative treatment effects between active interventions (appendix p 228). Overall, 46 (9%) of 522 trials were rated as high risk of bias, 380 (73%) trials as moderate, and 96 (18%) as low (appendix pp 115–32).

We also synthesised head-to-head studies separately to assess the differences between drugs. Figure 4 presents these data for the primary outcomes (194 studies with at least two active groups at licensed dose and

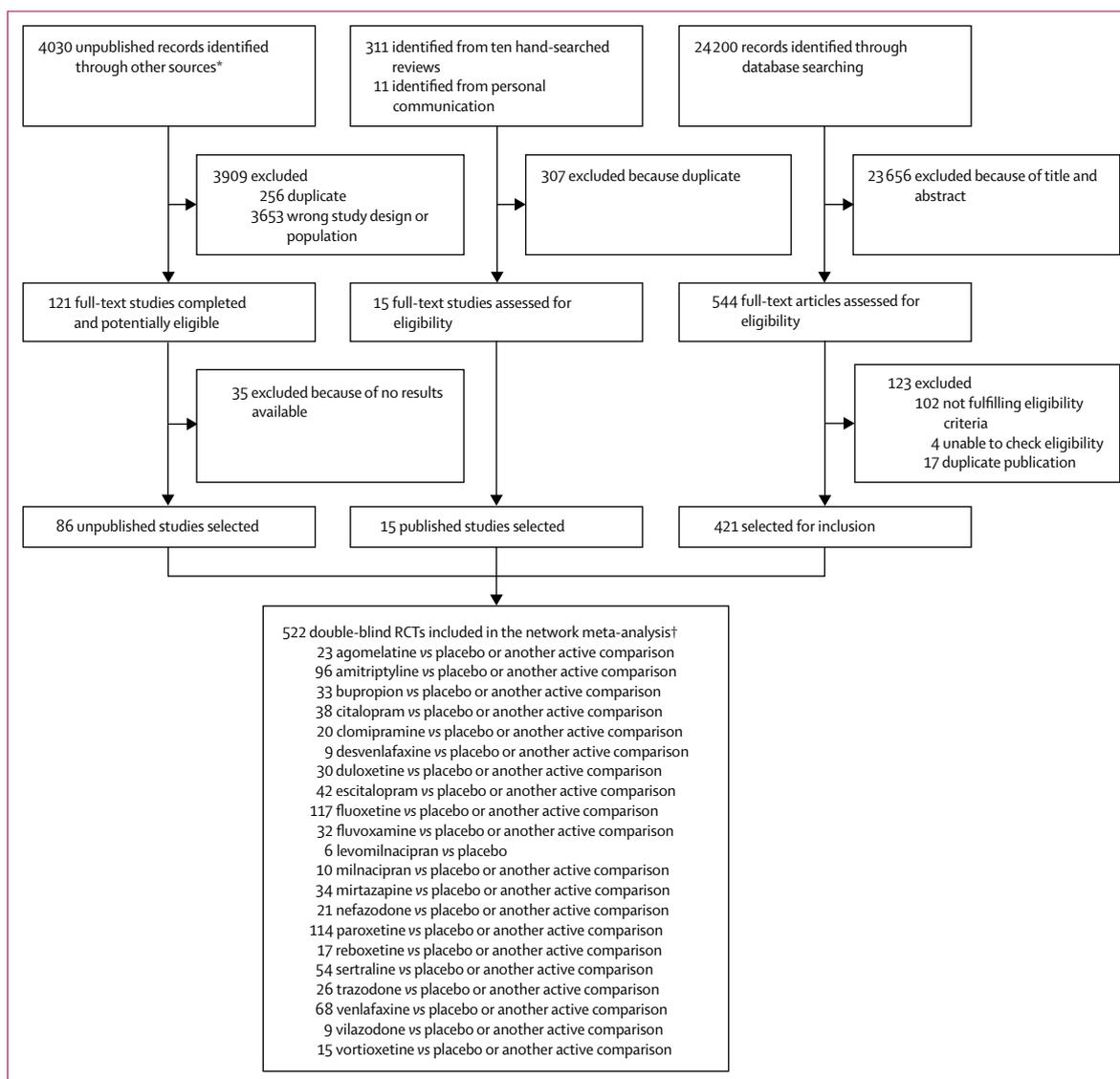


Figure 1: Study selection process

RCTs=randomised controlled trials. *Industry websites, contact with authors, and trial registries. The total number of unpublished records is the total number of results for each drug and on each unpublished database source. †522 RCTs corresponded to 814 treatment groups.

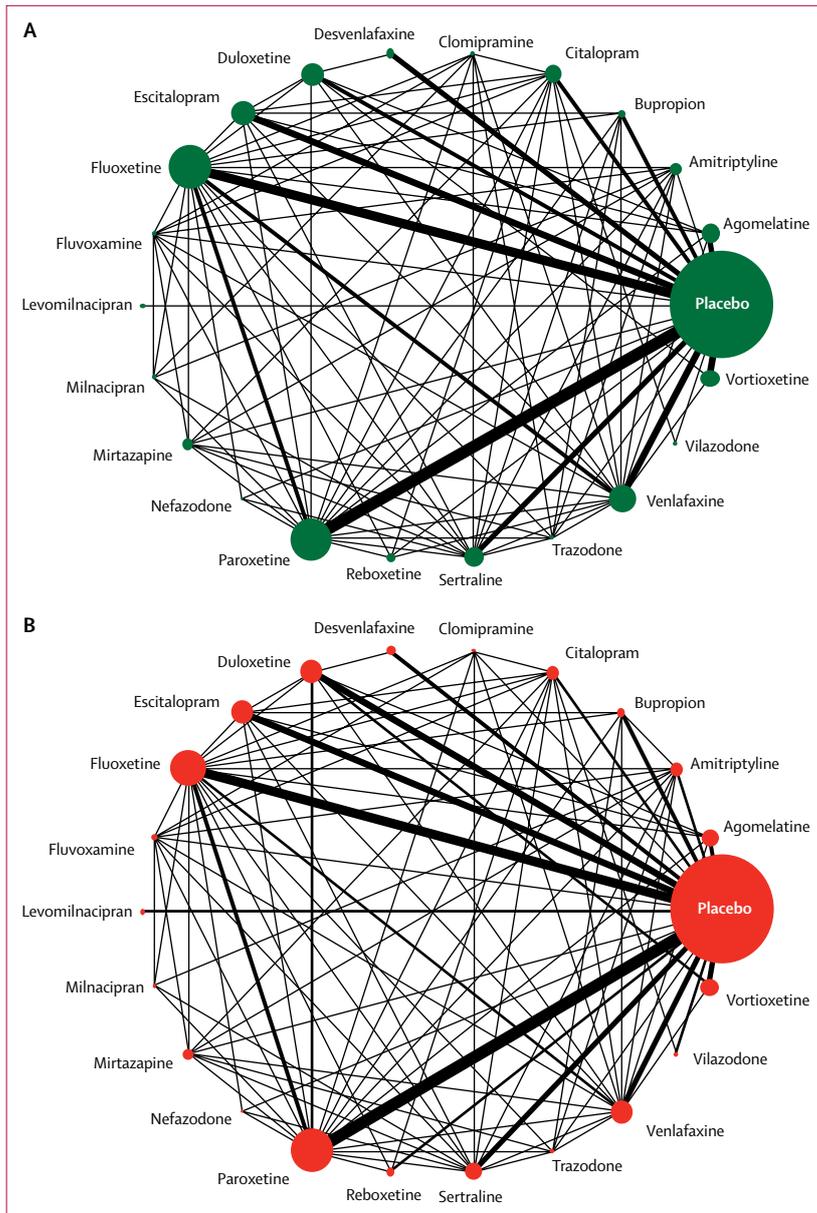


Figure 2: Network meta-analysis of eligible comparisons for efficacy (A) and acceptability (B)
Width of the lines is proportional to the number of trials comparing every pair of treatments. Size of every circle is proportional to the number of randomly assigned participants (ie, sample size).

comprised 34 196 patients). Agomelatine, amitriptyline, escitalopram, mirtazapine, paroxetine, venlafaxine, and vortioxetine were more effective than other antidepressants (ORs ranging between 1.19 and 1.96), whereas fluoxetine, fluvoxamine, reboxetine, and trazodone were among the least efficacious drugs (ORs ranging between 0.51 and 0.84). In terms of acceptability, agomelatine, citalopram, escitalopram, fluoxetine, sertraline, and vortioxetine were more tolerable than other antidepressants (ORs ranging between 0.43 and 0.77), whereas amitriptyline,

clomipramine, duloxetine, fluvoxamine, reboxetine, trazodone, and venlafaxine were the antidepressants associated with the highest dropout rates (ORs ranging between 1.30 and 2.32). When all trials were considered, differences in ORs between antidepressants ranged from 1.15 to 1.55 for efficacy (appendix p 147) and from 0.64 to 0.83 for acceptability (appendix p 149), with wide CrIs on most of the comparative analyses. Figure 5 reports the two-dimensional graphs about efficacy and acceptability in all studies and head-to-head studies. Results for the secondary outcomes were in line with the findings for the primary outcomes (appendix pp 197–230). Within the head-to-head comparisons, when a treatment was the novel or experimental drug of comparison, it appeared to be significantly more effective than when that same treatment was the older or control drug of comparison (difference 1.18-times, 95% CrI 1.09–1.27). Adjusting for this novelty effect diminished the differences between antidepressants.

We incorporated the GRADE judgments in figure 4. The certainty of evidence for the relative treatment effects of efficacy and acceptability varied; it was moderate for most of the comparisons involving agomelatine, escitalopram, citalopram, and mirtazapine, and low to very low for most comparisons involving vortioxetine, nefazodone, clomipramine, bupropion, and amitriptyline (appendix pp 231–65). The appendix (pp 266–85) presents the ranking of treatments based on cumulative probability plots and SUCRAs.

In accordance with the review protocol, we also did a sensitivity analysis, including all the studies that used the drugs within the accepted doses (ie, doses recommended in some international clinical guidelines; appendix pp 133, 134) and the results did not change substantially (appendix p 187).

Discussion

This study is based on 522 double-blind studies, which included 116 477 patients randomly assigned to 21 individual first-generation and second-generation antidepressant drugs or placebo. The project extends our previous work that had addressed 12 antidepressants with data for head-to-head comparisons.⁹ The present analysis is substantially more comprehensive because it includes 21 active treatments and placebo. The much larger evidence base (about 117 000 vs 26 000 patients), obtained through exhaustive search for published and unpublished information, allowed us to investigate additional important outcomes, such as remission, change in mood symptoms and dropouts due to side-effects, and a number of methodological issues, such as sponsorship, dosing schedule, study precision, and novelty effect.¹⁶

We found that all antidepressants included in the meta-analysis were more efficacious than placebo in adults with major depressive disorder and the summary effect sizes were mostly modest. Some antidepressants, such as

escitalopram, mirtazapine, paroxetine, agomelatine, and sertraline had a relatively higher response and lower dropout rate than the other antidepressants. By contrast, reboxetine, trazodone, and fluvoxamine were associated with generally inferior efficacy and acceptability profiles compared with the other antidepressants, making them less favourable options. To make our results as relevant and robust as possible to inform clinical practice, we decided to focus on head-to-head studies and at the same time emphasise the certainty of the retrieved evidence. Our assessment overall found few differences between antidepressants when all data were considered, while there was more diversity in the range of efficacy and dropout patterns seen across the head-to-head comparisons than the meta-analysis of antidepressants versus placebo.

The present findings in adults contrast with the efficacy of antidepressants in children and adolescents, for which fluoxetine is probably the only antidepressant that might reduce depressive symptoms.²¹ This differential efficacy across age groups might reflect heterogeneous mechanisms and causes of depression,²² smaller number of studies in young people, or different methodological issues affecting adult and paediatric trials.²³ The effect sizes were also smaller in more recent and larger placebo-controlled trials than in older and smaller ones, which might be an indicator of bias.

Estimated differences between drugs were smaller in placebo-controlled trials than in head-to-head studies. There are several potential explanations, as many factors have been associated with higher placebo response rates, such as randomisation ratio and the expectation of receiving an active treatment, the therapeutic setting, or the frequency of study visits.²⁴ In our dataset, we found that response to the same antidepressant was on average smaller and dropouts more likely to occur in placebo controlled trials than in head-to-head studies. Moreover, for the same drug and the same probability of receiving placebo, larger all-cause dropout rates were associated with a lower response to treatment. The use of the last observation carried forward (LOCF) approach for imputing missing outcome data might have affected the estimates of treatment effect.²⁵ Depressive symptoms tend to spontaneously improve over time and this phenomenon contributes to the high percentage of placebo responders in antidepressant trials.²⁶ Patients randomly assigned to the active drug in a double-blind, placebo-controlled trial might leave studies earlier than in head-to-head studies because they might suspect they have been allocated to the placebo group than to the intervention group. Antidepressants usually take full effect only after weeks of treatment; therefore, participants who dropped out earlier tend to have poorer responses than those who remain on treatment, which are carried forward to the end of the trial by the LOCF analysis. The final result can be an underestimate of the true efficacy of the active drug.

Another possible explanation could be a bias in conduct, analysis, or reporting of head-to-head trials, driven by

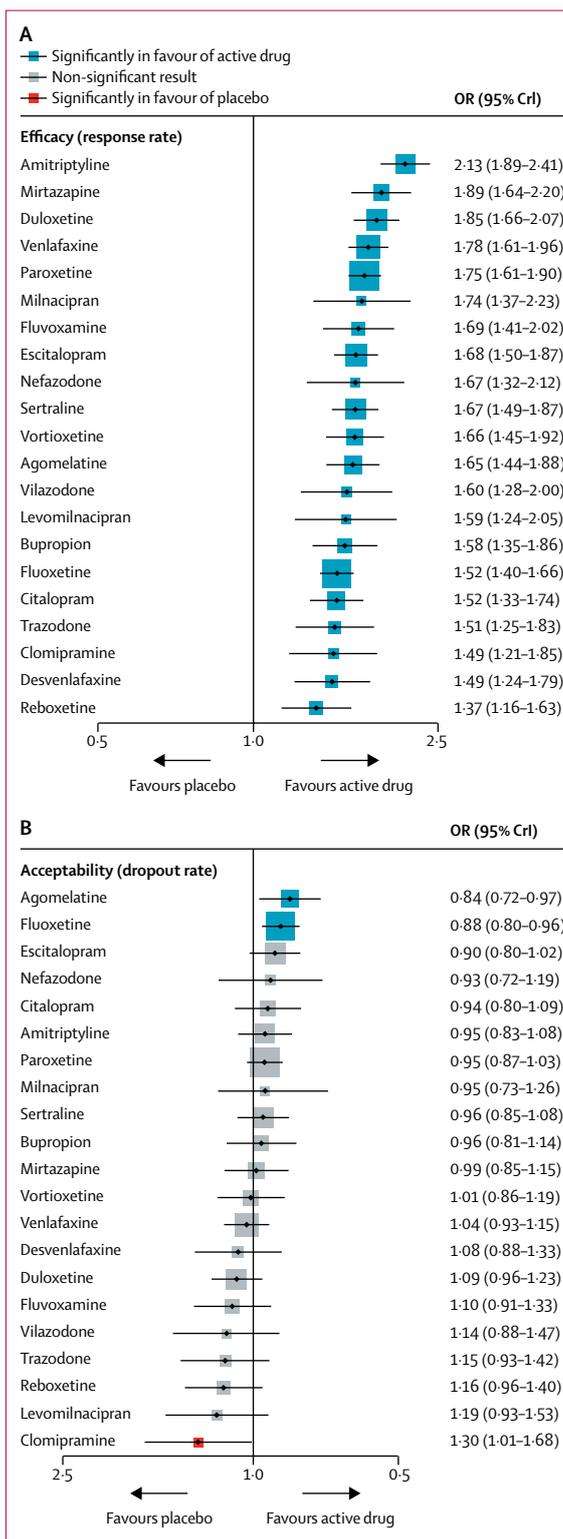


Figure 3: Forest plots of network meta-analysis of all trials for efficacy (A) and acceptability (B)

Antidepressants were compared with placebo, which was the reference compound. OR=odds ratio. CrI=credible interval.

□ Efficacy (response rate) ■ Comparison □ Acceptability (dropout rate)

Agom	0.72* (0.55-0.92)	0.80* (0.54-1.15)	0.89* (0.66-1.19)	0.57* (0.42-0.77)	0.62† (0.47-0.82)	0.97* (0.74-1.27)	0.85† (0.68-1.05)	0.69† (0.51-0.97)	0.79* (0.58-1.09)	0.81* (0.61-1.05)	0.70* (0.44-1.14)	0.81* (0.65-1.00)	0.53* (0.36-0.80)	0.86* (0.66-1.13)	0.69* (0.48-0.98)	0.74† (0.58-0.92)	1.24† (0.71-2.19)
0.96* (0.76-1.24)	Amit	1.10† (0.78-1.58)	1.23* (0.94-1.64)	0.79† (0.60-1.05)	0.87† (0.66-1.15)	1.35* (1.05-1.74)	1.18† (0.99-1.42)	0.97† (0.74-1.24)	1.10† (0.84-1.45)	1.12* (0.89-1.42)	0.98† (0.62-1.55)	1.12† (0.95-1.34)	0.74† (0.51-1.10)	1.20* (0.97-1.47)	0.96† (0.70-1.31)	1.02† (0.83-1.26)	1.72† (1.00-3.05)
0.87† (0.59-1.30)	0.91† (0.62-1.31)	Bupr	1.11† (0.76-1.67)	0.71† (0.49-1.07)	0.78† (0.53-1.18)	1.23* (0.84-1.80)	1.07† (0.76-1.50)	0.87† (0.59-1.30)	1.00† (0.66-1.49)	1.01† (0.70-1.47)	0.89† (0.51-1.54)	1.02† (0.73-1.43)	0.67† (0.42-1.08)	1.08† (0.75-1.56)	0.87† (0.57-1.30)	0.92† (0.66-1.30)	1.55† (0.85-2.94)
1.13* (0.88-1.47)	1.18* (0.93-1.49)	1.30† (0.88-1.93)	Cita	0.64† (0.47-0.87)	0.70* (0.51-0.95)	1.09* (0.85-1.42)	0.96* (0.76-1.21)	0.78* (0.57-1.06)	0.89* (0.64-1.23)	0.91† (0.68-1.21)	0.79† (0.49-1.32)	0.91* (0.71-1.17)	0.60† (0.41-0.87)	0.97† (0.74-1.25)	0.77* (0.53-1.13)	0.83† (0.64-1.07)	1.40† (0.78-2.48)
1.20* (0.91-1.59)	1.24† (0.98-1.58)	1.37† (0.93-2.04)	1.06* (0.82-1.38)	Clom	1.10† (0.80-1.51)	1.71* (1.27-2.29)	1.49† (1.16-1.99)	1.22† (0.88-1.67)	1.40† (1.00-1.92)	1.41* (1.05-1.91)	1.24† (0.76-2.00)	1.42† (1.12-1.79)	0.94† (0.62-1.41)	1.51† (1.15-1.96)	1.21† (0.83-1.73)	1.29† (0.99-1.67)	2.20† (1.22-3.90)
1.06* (0.82-1.37)	1.10† (0.84-1.42)	1.21† (0.81-1.81)	0.93* (0.71-1.22)	0.88† (0.66-1.18)	Dulo	1.56* (1.19-2.01)	1.37* (1.06-1.73)	1.12* (0.80-1.53)	1.28† (0.91-1.75)	1.30* (0.96-1.72)	1.13† (0.69-1.83)	1.30* (1.02-1.63)	0.86† (0.57-1.29)	1.38† (1.04-1.80)	1.10† (0.76-1.59)	1.18† (0.92-1.49)	1.99† (1.13-3.52)
0.90* (0.71-1.14)	0.93* (0.74-1.17)	1.03† (0.70-1.51)	0.79* (0.55-0.97)	0.75* (0.58-0.97)	0.85* (0.67-1.08)	Esci	0.70* (0.57-0.87)	0.71* (0.53-0.96)	0.81* (0.60-1.11)	0.83* (0.63-1.08)	0.72† (0.45-1.18)	0.83* (0.67-1.03)	0.55* (0.37-0.81)	0.88* (0.69-1.12)	0.70* (0.49-1.00)	0.75* (0.60-0.94)	1.27† (0.73-2.25)
1.20* (0.99-1.48)	1.25† (1.06-1.48)	1.38† (0.97-1.97)	1.06* (0.87-1.29)	1.00† (0.81-1.24)	1.14* (0.91-1.44)	1.34* (1.11-1.61)	Fluo	0.82* (0.64-1.04)	0.94* (0.72-1.20)	0.95* (0.77-1.16)	0.83† (0.54-1.30)	0.95* (0.83-1.09)	0.63† (0.44-0.90)	1.01† (0.84-1.21)	0.81* (0.60-1.09)	0.87† (0.74-1.01)	1.46† (0.85-2.53)
1.20* (0.91-1.61)	1.25† (0.99-1.59)	1.38† (0.93-2.07)	1.06* (0.82-1.39)	1.01† (0.76-1.32)	1.14† (0.85-1.54)	1.34* (1.03-1.75)	1.00* (0.80-1.25)	Fluv	1.14† (0.84-1.56)	1.16* (0.89-1.52)	1.01† (0.62-1.71)	1.16* (0.90-1.49)	0.77† (0.51-1.17)	1.23* (0.94-1.63)	0.99† (0.69-1.42)	1.06* (0.80-1.38)	1.78† (1.00-3.24)
1.07* (0.80-1.44)	1.11† (0.86-1.43)	1.23† (0.81-1.85)	0.94† (0.71-1.26)	0.89† (0.67-1.19)	1.01† (0.74-1.38)	1.19* (0.90-1.58)	0.89* (0.70-1.13)	0.89† (0.67-1.17)	Miln	1.02† (0.75-1.37)	0.88† (0.54-1.44)	1.02† (0.80-1.31)	0.67† (0.45-1.03)	1.08* (0.82-1.44)	0.86* (0.60-1.25)	0.93* (0.71-1.22)	1.56† (0.89-2.84)
0.93* (0.72-1.21)	0.97* (0.77-1.21)	1.07† (0.73-1.57)	0.82* (0.65-1.05)	0.78* (0.60-1.01)	0.88* (0.67-1.16)	1.04* (0.82-1.32)	0.78* (0.64-0.94)	0.78* (0.60-0.99)	0.87* (0.66-1.15)	Mirt	0.87† (0.55-1.41)	1.00* (0.82-1.23)	0.66* (0.45-0.99)	1.06* (0.84-1.35)	0.85* (0.62-1.18)	0.91* (0.73-1.13)	1.53† (0.89-2.72)
1.15† (0.76-1.76)	1.19† (0.80-1.78)	1.32† (0.80-2.20)	1.01† (0.67-1.54)	0.96† (0.63-1.45)	1.09† (0.71-1.68)	1.28* (0.86-1.94)	0.96† (0.66-1.40)	0.95† (0.63-1.46)	1.07† (0.70-1.67)	1.23* (0.82-1.86)	Nefa	1.15† (0.74-1.78)	0.75† (0.43-1.32)	1.23† (0.77-1.90)	0.98† (0.57-1.64)	1.04† (0.66-1.65)	1.76† (0.90-3.56)
1.01* (0.82-1.24)	1.05† (0.89-1.23)	1.16† (0.81-1.64)	0.89* (0.72-1.09)	0.84* (0.68-1.03)	0.95† (0.76-1.19)	1.12* (0.93-1.35)	0.84* (0.73-0.95)	0.84* (0.67-1.04)	0.94† (0.75-1.18)	1.08* (0.89-1.30)	0.88† (0.60-1.27)	Paro	0.66† (0.46-0.94)	1.06* (0.88-1.28)	0.85† (0.63-1.15)	0.91* (0.77-1.07)	1.53† (0.90-2.66)
1.44* (1.02-2.04)	1.50† (1.07-2.07)	1.65† (1.05-2.60)	1.27† (0.92-1.75)	1.20† (0.84-1.70)	1.36† (0.95-1.95)	1.60* (1.14-2.23)	1.20† (0.88-1.62)	1.20† (0.83-1.71)	1.35† (0.92-1.95)	1.54* (1.09-2.17)	1.25† (0.77-2.01)	1.43† (1.05-1.94)	Rebo	1.61† (1.09-2.34)	1.29† (0.81-2.01)	1.38† (0.94-1.99)	2.32† (1.24-4.41)
1.07* (0.85-1.37)	1.11* (0.92-1.35)	1.23† (0.85-1.79)	0.95† (0.76-1.18)	0.90† (0.71-1.13)	1.02† (0.79-1.32)	1.20* (0.97-1.48)	0.89† (0.76-1.05)	0.89† (0.70-1.13)	1.00† (0.77-1.30)	1.15* (0.93-1.43)	0.93† (0.63-1.37)	1.07* (0.90-1.26)	0.75† (0.54-1.04)	Sert	0.80* (0.58-1.11)	0.86* (0.70-1.05)	1.45† (0.84-2.54)
1.36* (0.99-1.87)	1.41† (1.06-1.86)	1.56† (1.04-2.31)	1.20* (0.88-1.63)	1.13† (0.83-1.54)	1.28† (0.92-1.79)	1.51* (1.12-2.04)	1.13† (0.87-1.46)	1.13† (0.82-1.55)	1.27* (0.91-1.76)	1.45* (1.09-1.94)	1.18† (0.75-1.84)	1.35* (1.04-1.75)	0.94† (0.64-1.39)	1.26† (0.95-1.67)	Traz	1.07† (0.77-1.47)	1.80† (0.98-3.38)
1.01* (0.82-1.26)	1.05† (0.87-1.27)	1.16† (0.82-1.65)	0.90† (0.72-1.10)	0.85† (0.67-1.06)	0.96† (0.77-1.21)	1.13* (0.93-1.37)	0.84† (0.73-0.97)	0.84* (0.66-1.07)	0.95* (0.73-1.23)	1.09* (0.89-1.33)	0.88† (0.59-1.30)	1.01† (0.86-1.17)	0.70† (0.51-0.97)	0.94* (0.78-1.13)	0.75† (0.57-0.98)	Venl	1.69† (1.01-2.86)
0.73† (0.42-1.26)	0.76† (0.44-1.29)	0.83† (0.45-1.54)	0.64† (0.37-1.11)	0.61† (0.35-1.05)	0.69† (0.40-1.20)	0.81† (0.47-1.39)	0.60† (0.36-1.02)	0.60† (0.34-1.05)	0.68† (0.39-1.20)	0.78† (0.45-1.34)	0.63† (0.33-1.19)	0.72† (0.43-1.22)	0.51† (0.28-0.92)	0.68† (0.39-1.16)	0.54† (0.30-0.95)	0.72† (0.43-1.19)	Vort

Figure 4: Head-to-head comparisons for efficacy and acceptability of the 21 antidepressants
 Drugs are reported in alphabetical order. Data are ORs (95% CrI) in the column-defining treatment compared with the row-defining treatment. For efficacy, ORs higher than 1 favour the column-defining treatment (ie, the first in alphabetical order). For acceptability, ORs lower than 1 favour the first drug in alphabetical order. To obtain ORs for comparisons in the opposite direction, reciprocals should be taken. Significant results are in bold and underscored. The certainty of the evidence (according to GRADE) was incorporated in this figure (appendix pp 231-65). OR=odds ratio. CrI=credible interval. Agom=agomelatine. Amit=amitriptyline. Bupr=bupropion. Cita=citalopram. Clom=clomipramine. Dulo=duloxetine. Esci=escitalopram. Fluo=fluoxetine. Fluv=fluvoxamine. Miln=milnacipran. Mirt=mirtazapine. Nefa=nefazodone. Paro=paroxetine. Rebo=reboxetine. Sert=sertraline. Traz=trazodone. Venl=venlafaxine. Vort=vortioxetine. *Moderate quality of evidence. †Low quality of evidence. ‡Very low quality of evidence.

commercial interests.²⁷ In our analyses, funding by industry was not associated with substantial differences in terms of response or dropout rates. However, non-industry funded trials were few and many trials did not report or disclose any funding. We also observed that drugs tended to show a better efficacy profile when they were novel and used as experimental treatments than when they had become old. This novelty effect might arise where a novel agent is perceived to be more effective and better tolerated; alternatively, selective analyses and outcome reporting bias might be more prominent when a treatment is first launched.¹⁶

Our literature search was as comprehensive as possible, including the largest amount of unpublished data to date, which are associated with less favourable effect sizes for antidepressants.²⁸ Although it is possible that a certain amount of unpublished data could not be retrieved, our comparison-adjusted funnel plots did not suggest that small studies gave different results from larger studies

either among placebo-controlled trials or head-to-head comparison trials (appendix pp 179-81, 225-27). The estimates of treatment effect from our study are in line with previous reviews on the same matter,²⁸ but they are considerably more precise because of our larger quantity of data and resulting statistical power.

Our review has some limitations. According to the GRADE framework, the quality of many comparisons was assessed as low or very low for amitriptyline, bupropion, and venlafaxine, whereas it was often rated as moderate for agomelatine, escitalopram, and mirtazapine. We incorporated the certainty of evidence in the main results of our analysis to highlight the most robust findings for further use in clinical judgment. However, many trials did not report adequate information about randomisation and allocation concealment, which restricts the interpretation of these results. To increase the methodological rigour of the contributing evidence, we included only double-blind trials, which were generally very similar in design and

conduct. The poor information in terms of risk of bias assessment might be a matter of reporting; however, we presented full details about the risk of bias of all included studies in the appendix (pp 115–32). We did not do a formal cost-effectiveness analysis. All of the most effective antidepressants are now off patent and available in generic form. Some of the antidepressants are included in the WHO Model List of Essential Medicines, which makes them available worldwide and ready to use also in developing countries.

We analysed only average treatment effects and were not able to investigate potentially important clinical and demographical modifiers of treatment response at the individual patient level (eg, age, sex, severity of symptoms, or duration of illness). Patients recruited in randomised trials tend to be highly selected and we also excluded patients with psychotic or treatment-resistant depression, which might limit the applicability of the results to these clinical subgroups, but it was intended as a methodological strength to assure transitivity in the network. We did not cover important clinical issues that might inform treatment decision making in routine clinical practice (eg, specific adverse events, withdrawal symptoms, or combination with non-pharmacological treatments). Additionally, because of the paucity of information reported in the original studies, we were not able to quantify some outcomes, such as global functioning. It should also be noted that some of the adverse effects of antidepressants occur over a prolonged period, meaning that positive results need to be taken with great caution, because the trials in this network meta-analysis were of short duration. The current report summarises evidence of differences between antidepressants when prescribed as an initial treatment. Given the modest effect sizes, non-response to antidepressants will occur. Our information unfortunately cannot guide next-step choices after failure of such a first step (ie, they do not apply to treatment-resistant depression), for which well performed trials are scarce.²⁹

Using the data made available on the websites of the US Food and Drug Administration and European Medicines Agency, on the international trial registries, and from contacting study authors and pharmaceutical companies, we managed to incorporate in the analysis a considerable amount of unpublished data for some drugs—namely, agomelatine, escitalopram, paroxetine, reboxetine, sertraline, venlafaxine, vilazodone, and vortioxetine—but not for all the antidepressants included in the network meta-analysis. This limitation in the primary trials might affect the validity of the findings for some antidepressants, but the incorporation of both direct and indirect comparisons might have contributed to reduce the potential risk of bias.³⁰ We did our best to retrieve all unpublished data and contacted study authors for supplemental material, but we are aware that a substantial amount of information is still not available to the public. There are online archives where trials are

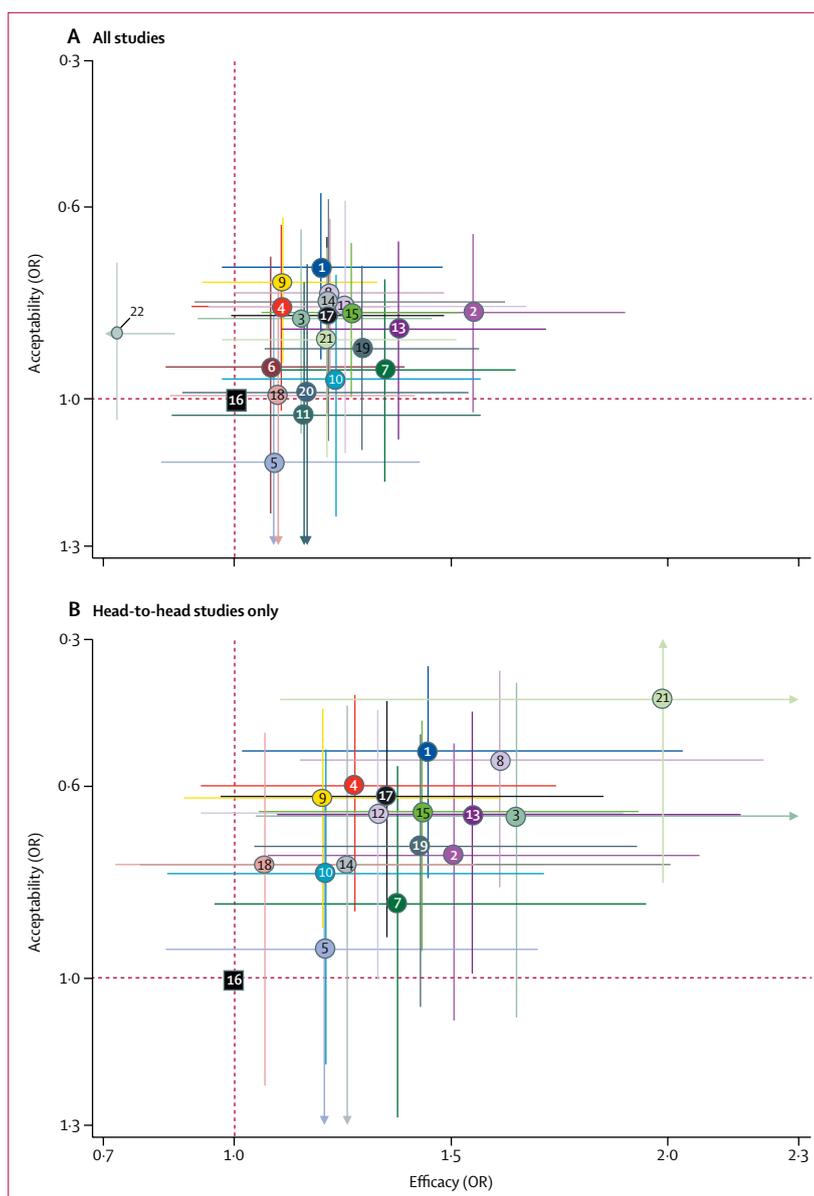


Figure 5: Two-dimensional graphs about efficacy and acceptability in all studies (A) and head-to-head (B) studies only

Data are reported as ORs in comparison with reboxetine, which is the reference drug. Error bars are 95% CrIs. Individual drugs are represented by different coloured nodes. Desvenlafaxine, levomilnacipran, and vilazodone were not included in the head-to-head analysis because these three antidepressants had only placebo-controlled trials. ORs=odds ratios. 1=agomelatine. 2=amitriptyline. 3=bupropion. 4=citalopram. 5=clomipramine. 6=desvenlafaxine. 7=duloxetine. 8=escitalopram. 9=fluoxetine. 10=fluvoxamine. 11=levomilnacipran. 12=milnacipran. 13=mirtazapine. 14=nefazodone. 15=paroxetine. 16=reboxetine. 17=sertraline. 18=trazodone. 19=venlafaxine. 20=vilazodone. 21=vortioxetine. 22=placebo.

prospectively registered; however, they collect reliable information only about the most recent studies and we cannot rule out the possibility that some studies are absent or the same study has been counted twice in our analyses. It is not uncommon for the same study to go by different names in different publications, which complicates the process of data synthesis.³¹ By making the dataset fully and freely available, we welcome any

information that might help clarify any mistakes in our dataset.

Notwithstanding these limitations, the findings from this network meta-analysis represent the most comprehensive currently available evidence base to guide the initial choice about pharmacological treatment for acute major depressive disorder in adults. All statements comparing the merits of one antidepressant with another must be tempered by the potential limitations of the methodology,³² the complexity of specific patient populations, and the uncertainties that might result from choice of dose or treatment setting. We hope that these results will assist in shared decision making between patients, carers, and their clinicians.

Contributors

ACi, TAF, GS, and JRG conceived and designed the study. ACi, TAF, LZA, SL, HGR, YO, NT, YH, EHT, HI, KS, and AT selected the articles and extracted the data. GS, Ach, JPTH, and ME analysed the data. ACi, TAF, GS, and JRG wrote the first draft of the manuscript. Ach, LZA, YO, SL, HGR, EHT, JPTH, ME, and JPAI interpreted the data and contributed to the writing of the final version of the manuscript. All authors agreed with the results and conclusions of this Article.

Declaration of interests

ACi is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility. TAF has received lecture fees from Eli Lilly, Janssen, Meiji, Mitsubishi-Tanabe, Merck Sharp & Dohme, and Pfizer; consultancy fees from Takeda Science Foundation; and research support from Mochida and Mitsubishi-Tanabe. SL has received honoraria for consulting from LB Pharma, Lundbeck, Otsuka, TEVA, Geodon Richter, Recordati, LTS Lohmann, and Boehringer Ingelheim; and for lectures from Janssen, Lilly, Lundbeck, Otsuka, SanofiAventis, and Servier. NT has received lecture fees from Otsuka and Meiji. YH has received lecture fees from Yoshitomi. JRG is an NIHR Senior Investigator. All other authors declare no competing interests.

Acknowledgments

This study was funded by the National Institute for Health Research (NIHR) Oxford Health Biomedical Research Centre (BRC-1215-20005) and the Japan Society for the Promotion of Science (17K19808). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health.

References

- 1 GBD 2013 DALYs and HALE Collaborators, Murray CJ, Barber RM, et al. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. *Lancet* 2015; **386**: 2145–91.
- 2 GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388**: 1603–58.
- 3 WHO. Depression: fact sheet. Geneva: World Health Organisation, 2017. <http://www.who.int/mediacentre/factsheets/fs369/en/> (accessed Sept 21, 2017).
- 4 Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008; **3**: 14.
- 5 Harmer CJ, Duman RS, Cowen PJ. How do antidepressants work? New perspectives for refining future treatment approaches. *Lancet Psychiatry* 2017; **4**: 409–18.
- 6 Higgins JP, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? *Lancet* 2015; **386**: 628–30.
- 7 Hróbjartsson A, Thomsen AS, Emanuelsson F, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. *CMAJ* 2013; **185**: E201–11.
- 8 Furukawa TA, Salanti G, Atkinson LZ, et al. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. *BMJ Open* 2016; **6**: e010919.
- 9 Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatment meta-analysis. *Lancet* 2009; **373**: 746–58.
- 10 Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. *Int Clin Psychopharmacol* 2005; **20**: 49–52.
- 11 Rhodes KM, Turner RM, Higgins JP. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. *J Clin Epidemiol* 2015; **68**: 52–60.
- 12 Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. *J Clin Epidemiol* 2011; **64**: 163–71.
- 13 Higgins JP, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods* 2012; **3**: 98–110.
- 14 Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. *Stat Med* 2010; **29**: 932–44.
- 15 Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP. Evaluating the quality of evidence from a network meta-analysis. *PLoS One* 2014; **9**: e99682.
- 16 Salanti G, Dias S, Welton NJ, et al. Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med* 2010; **29**: 2369–83.
- 17 Chaimani A, Higgins JP, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. *PLoS One* 2013; **8**: e76654.
- 18 Lunn D, Spiegelhalter D, Thomas A, et al. The BUGS project: evolution, critique and future directions. *Stat Med* 2009; **28**: 3049–67.
- 19 Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: the network graphs package. *Stata J* 2015; **15**: 905–50.
- 20 Schwarzer G. Network meta-analysis. In: Schwarzer G, Carpenter JR, Rücker G, eds. *Meta-analysis with R*. Berlin: Springer, 2015: 187–216.
- 21 Cipriani A, Zhou X, Del Giovane C, et al. Comparative efficacy and tolerability of antidepressants for major depressive disorder in children and adolescents: a network meta-analysis. *Lancet* 2016; **388**: 881–90.
- 22 Thapar A, Collishaw S, Pine DS, Thapar AK. Depression in adolescence. *Lancet* 2012; **379**: 1056–67.
- 23 Walkup JT. Antidepressant efficacy for depression in children and adolescents: industry- and NIMH-funded studies. *Am J Psychiatry* 2017; **174**: 430–37.
- 24 Rutherford BR, Roose SP. A model of placebo response in antidepressant clinical trials. *Am J Psychiatry* 2013; **170**: 723–33.
- 25 Cook RJ, Zeng L, Yi GY. Marginal analysis of incomplete longitudinal binary data: a cautionary note on LOCF imputation. *Biometrics* 2004; **60**: 820–28.
- 26 Furukawa TA, Cipriani A, Atkinson LZ, et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. *Lancet Psychiatry* 2016; **3**: 1059–66.
- 27 Perlis RH, Perlis CS, Wu Y, Hwang C, Joseph M, Nierenberg AA. Industry sponsorship and financial conflict of interest in the reporting of clinical trials in psychiatry. *Am J Psychiatry* 2005; **162**: 1957–60.
- 28 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358**: 252–60.
- 29 Furukawa TA, Akechi T, Shimodera S, et al. Strategic use of new generation antidepressants for depression: SUN(^_^)D study protocol. *Trials* 2011; **12**: 116.
- 30 Furukawa TA, Miura T, Chaimani A, et al. Using the contribution matrix to evaluate complex study limitations in a network meta-analysis: a case study of bipolar maintenance pharmacotherapy review. *BMC Res Notes* 2016; **9**: 218.
- 31 Wager E. The need for trial identifiers. *Curr Med Res Opin* 2004; **20**: 203–06.
- 32 Ioannidis JP. Meta-analyses can be credible and useful: a new standard. *JAMA Psychiatry* 2017; **74**: 311–12.