

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Flaws in the evidence base of antidepressants for depression: A reanalysis and critical review of a network meta-analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-024886
Article Type:	Research
Date Submitted by the Author:	19-Jun-2018
Complete List of Authors:	Munkholm, Klaus; Nordic Cochrane Centre, Righospitalet, 7811 Paludan-Müller, Asger; Nordic Cochrane Centre, Righospitalet, 7811 Boesen, Kim; Nordic Cochrane Centre, Righospitalet, 7811
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

Peer Review Only

1
2
3
4
5 **Flaws in the evidence base of antidepressants for depression: A**
6 **reanalysis and critical review of a network meta-analysis**
7
8
9
10

11 **Authors**

12 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a
13
14

15 **Author affiliations**

16 ^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
17
18
19
20
21
22
23

24 **Word count:** 4440 words.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 **Correspondence to:**

45 Klaus Munkholm, M.D., DMSc.
46 Nordic Cochrane Centre
47 Rigshospitalet, 7811
48 Blegdamsvej 9
49 2100 København Ø
50 Denmark
51 Phone: +45 35457112
52 E-mail: km@cochrane.dk
53
54
55
56
57
58
59
60

Abstract

Objectives

To investigate whether the conclusion of a recent systematic review and network meta-analysis (Cipriani et al. (2018)) that antidepressants are more efficacious than placebo for adult depression was supported by the evidence.

Design

Reanalysis of systematic review, with meta-analysis.

Data sources

522 trials (116 477 participants) as reported in the systematic review by Cipriani et al. (2018) and clinical study reports for 19 of these trials.

Analysis

The Cochrane Handbook's risk of bias tool and the GRADE approach were used to evaluate the risk of bias and the certainty of evidence, respectively. The impact of the use of a "placebo run-in" study design and publication status was estimated using direct pair-wise meta-analyses.

Results

Many flaws in the evidence base, which lead to inflation of the effect size for antidepressants, were unrecognised or underestimated by Cipriani et al. We found higher effect sizes for antidepressants versus placebo in trials with a "placebo run-in" study design (SMD of 0.31 (95% confidence interval 0.28 to 0.34)) compared to trials without a "placebo run-in" design (SMD of 0.22 (0.16 to 0.28)) and lower effect sizes in unpublished studies (SMD 0.15 (0.11 to 0.19)) compared with published studies (SMD 0.33 (0.30 to 0.35)). The outcome data reported by Cipriani et al. differed from the clinical study reports for 12 (63%) of the 19 trials we assessed. The certainty of evidence for placebo-controlled comparisons should be very low. The apparent effect of antidepressants compared with placebo was not clinically relevant (MD of 1.97 points (1.74 to 2.21) on the 17-item Hamilton depression rating scale (range 0 to 52)), and the true effect may even be negative.

Conclusions

The conclusion of the systematic review by Cipriani et al (2018), that antidepressants are more efficacious than placebo for depression in adults, was not supported by the evidence.

Strengths and limitations of this study

- Empirical evidence was provided showing how a large number of flaws in the evidence base for antidepressants for depression lead to an inflation of the apparent effect size for antidepressants
- By performing pair-wise meta-analyses, we provided novel evidence regarding the association between the use of a “placebo run-in” trial design and the apparent effect size for antidepressants compared with placebo
- In pair-wise meta-analysis we provided an estimate of the effect of antidepressants compared with placebo on scores on the Hamilton depression rating scale, which was not available in the review by Cipriani et al.
- We compared the data reported by Cipriani et al. on the outcomes of total dropouts and dropouts due to adverse events with the clinical study reports that we previously obtained from the European Medicines Agency (EMA)
- Our review and meta-analyses relied on the data reported by Cipriani et al. and we did not perform a separate literature search and data extraction; given the flaws we have identified a more reliable assessment would need to be based on the original clinical study reports

Introduction

WHO estimates that 300 million people globally suffer from depression, making depression the leading cause of disability worldwide.¹ In Denmark, 10% of all adults 25 years and older were in treatment with antidepressants in 2016.² In the US, 13% of persons 12 years and older were in treatment in 2014, making antidepressants one of the three most commonly used drug classes.³ Prescriptions for antidepressants cost the National Health Service in the United Kingdom an estimated £267 million in 2016.⁴ Research that guides clinical treatment of depression therefore has a potentially important impact on millions of people and on national economies.

The recent network meta-analysis of antidepressants for depression by Cipriani et al.⁵ is the largest meta-analysis of antidepressants to date in terms of included studies and participants. It specifically aimed to inform clinical guidelines, patients, physicians and policy makers by comparing 21 antidepressants for the treatment of adults with depression. The review's primary outcomes were 'response rate' (defined as the number of participants with at least a 50% reduction on an observer-rated depression scale) and overall dropout rates. The secondary outcomes were depression symptom scores, 'remission rate' (defined as the number of participants with an observer-rated depression score below a certain threshold), and dropouts due to adverse events. Cipriani et al. found that all 21 antidepressants were more effective than placebo, whereas only two of the drugs had fewer dropouts compared with placebo. Based on these findings, they⁵ ranked the antidepressants according to 'response rate' and overall dropout rate and concluded that antidepressants were more efficacious than placebo in adults with major depressive disorder. The improvement in symptom scores they found were very similar to previous meta-analyses (Figure 1), some of which have concluded that the benefit of antidepressants is doubtful.⁶⁻⁹ The review received widespread media coverage, largely citing it as finally putting to rest any doubts regarding the efficacy of antidepressants,^{10 11} and the message of antidepressants being effective was strongly conveyed by some of the authors in the press,¹⁰ adding that the benefits outweigh side effects.¹¹

There are many methodological issues and flaws in trials of antidepressant agents,¹² of which many have been acknowledged for decades.¹³ Research aiming to inform clinical practice on the use of antidepressants for depression must recognise these limitations. Given the potential implications of Cipriani et al.'s review⁵ we therefore aimed to investigate how these flaws were addressed, whether the review's methods were appropriate and followed the protocol,¹⁴ and whether the conclusion was supported by the evidence. We furthermore aimed to provide empirical evidence on the impact of these methodological flaws by using the data reported by Cipriani et al.⁵

Methods

Data collection

We extracted the review's risk of bias assessments and descriptive data from the online supplement and converted the data to Microsoft Excel format. We downloaded the online dataset⁵ and merged the files for our statistical analyses.

We cross-referenced the included trials with the clinical study reports that we previously obtained from the European Medicines Agency (EMA) in 2010.¹⁵ We compared the outcomes of total dropout rates and dropouts due to adverse events as reported in the clinical study reports with the data reported by Cipriani et al.⁵

Statistical analyses

Descriptive analyses were made in Microsoft Excel. We used the statistical software R (version 3.4.3) for random effects meta-analyses based on the inverse variance method and calculated effect sizes as standardised mean differences (SMD) with corresponding 95% confidence intervals (95% CI). For the comparisons between antidepressants and placebo on rating scales, we used the Hartung-Knapp-Sidik-Jonkman approach because it results in fewer type I errors than the DerSimonian and Laird approach.¹⁶ We based our analyses on the number of participants from Cipriani et al.'s 'efficacy' analyses.⁵ In studies with more than one drug arm the total number of participants in the placebo group was split evenly between the active comparisons and the means and standard deviations were unchanged.¹⁷ We did subgroup analyses based on the use of a 'placebo run-in' study design and publication status, according to the study characteristics published by Cipriani et al.⁵

Quality assessments

We followed the Cochrane Handbook in our risk of bias assessments.¹⁷ We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE)¹⁸ approach to evaluate the certainty of evidence, which, for systematic reviews, reflects the extent of confidence that an estimate effect is correct. GRADE considers five domains that affect the quality of the evidence: the included trials' internal risk of bias, inconsistency of the included trials' results and large heterogeneity, indirectness of the evidence due to poor external validity, imprecision of the effect estimate and wide confidence intervals, and publication bias.¹⁸

Patient involvement

No patients were involved in the development of the research question, design and implementation of the study, or interpretation of the results.

Results

Risk of bias

We evaluated Cipriani et al.'s risk of bias assessments and whether they adhered to Cochrane's risk of bias tool.

Randomisation sequence generation and allocation concealment

Cipriani et al.⁵ judged 426 (82%) and 460 (88%) of the 522 included trials to be of unclear risk of bias with respect to randomisation sequence generation and allocation concealment, respectively. The remaining trials were of low risk of bias. Trials at high or unclear risk of bias within these two domains are likely to report inflated effect estimates, especially of subjective outcomes.¹⁹

Blinding of participants, personnel and investigators

Cipriani et al.⁵ did not use the standard Cochrane categorisation of low, unclear or high risk of bias due to a lack of blinding.¹⁷ They categorised instead 513 (98%) studies as 'stated-not tested' in at least one of the three blinding domains, meaning that the trial had stated to be double-blind, but did not test the blinding integrity. While this implied the presence of a blinding issue, their categorisation did not affect the overall risk of bias assessment⁵ and it seemed that the 'stated-not tested' domains were counted as 'low risk of bias'. Adverse effects of antidepressants are common and often reveal who receives active medication and who receives placebo in a randomised trial.

1
2
3
4 The degree of unblinding is extensive and leads to inflated effect estimates,²⁰ and smaller effects
5 have been observed when the trials were better blinded by adding atropine to the placebo.²¹ Two
6 of the three trials categorised by Cipriani et al.⁵ to be at low risk of bias in the blinding of
7 participants' domain had tested the blinding integrity (S1 Appendix). The blinding was likely
8 compromised in both trials. Cipriani et al.⁵ should arguably have categorised all placebo-controlled
9 trials as at least unclear, or perhaps even at high risk of bias.
10

11 **Selective outcome reporting**

12 Cipriani et al.⁵ judged that 402 (77%) of the 522 trials were of low risk of outcome reporting bias,
13 100 (19%) of unclear risk, and 20 (4%) of high risk of bias. Their assessments were based on the
14 reporting of the review's two primary outcomes of 'response rates' and overall dropout rates and
15 only if both outcomes were missing the trial was rated as high risk of bias. However, the review's
16 three secondary outcomes of dropouts due to adverse events, depression symptoms measured on
17 depression symptom scales, and 'remission rates' were not reported in 93 (18 %) trials, 98 (19%)
18 trials, and 71 (14 %) trials, respectively. In total, 182 (35%) trials did not report at least one primary
19 or secondary outcome and these trials should have been rated as high risk of bias.¹⁷ Selective
20 outcome reporting leads to overestimation of the benefits and underestimation of the harms of
21 interventions.²²
22
23
24

25 **Attrition bias**

26 Cipriani et al. rated trials that used an appropriate imputation method as low risk of bias.⁵ Trials
27 that used an 'inappropriate' imputation method were rated according to several arbitrary cut-offs:
28 When the dropout rates were unbalanced between the arms, defined as more than a 5% difference
29 for the head-to-head comparisons and a 10% difference for the placebo comparisons, they were
30 rated as high risk of bias. When the dropout rates between the arms were not unbalanced but the
31 total dropout rate was more than 20% they were rated unclear, and if the total dropout rate was
32 less than 20% they were rated as low risk of bias. This method is not in accordance with the
33 Cochrane Handbook. Furthermore, the authors did not consider the reasons for dropout, although
34 this is recommended by the Cochrane Handbook.¹⁷ The dropout rates due to adverse events were
35 higher for all 21 antidepressants than for placebo.⁵
36
37
38

39 According to Cipriani et al. 121 (23%) trials were at high risk of attrition bias, but we could not
40 replicate these results. The overall attrition rate was more than 20% in 334 (64%) trials. We found
41 that the dropout rates were unbalanced between the arms in 202 trials (39%) and they should have
42 been rated as high risk of bias unless an appropriate imputation method was used. The absolute
43 difference in dropout rates between the arms exceeded the authors' arbitrary cut-offs in 202 (39%)
44 trials and according to Cipriani et al.'s protocol, they should have been categorised at high risk of
45 bias unless an "appropriate imputation method" was used. Cipriani et al. characterised the last
46 observation carried forward (LOCF) method as inappropriate,¹⁴ however most antidepressant trials
47 use this imputation method.²³ The LOCF method may lead to an underestimation of the variability,
48 a falsely low p-value, and an overestimation of treatment effects.²⁴
49
50

51 **Other bias domain**

52 The authors omitted the 'other bias' domain from the risk of bias assessment although it is an
53 integrated part of Cochrane's risk of bias tool.¹⁷ Relevant biases were baseline imbalances and
54 design-specific risks of bias for cross-over and cluster randomised trials, which were eligible
55
56
57
58
59
60

1
2
3
4 according to the Cipriani et al. protocol,¹⁴ although the trial designs were not specified in the
5 review.⁵ Some argue that ‘vested interests’ should also be considered, since industry sponsored
6 drug studies lead to more favourable effects than other studies by mechanisms that are not
7 explained by the usual bias domains.²⁵
8

9 **Summary risk of bias assessment**

10 The authors deviated from Cochrane’s overall risk of bias categorisation of low, unclear or high risk
11 of bias,¹⁷ by introducing their own category of ‘moderate’ risk of bias.⁵ Also, they used a “vote
12 counting” system to categorise the trials, rather than an overall qualitative assessment as
13 recommended by the Cochrane Handbook.¹⁷ According to our knowledge, none of their methods
14 are supported by empirical evidence. The authors rated 96 (18%) of the 522 trials as low risk of
15 bias, 380 (73%) trials as ‘moderate’, and 46 (9%) trials as high risk of bias. We were not able to
16 replicate these findings. The review’s five outcomes were likely all affected by all the bias domains,
17 and trials with at least one or more ‘high risk of bias’ domains should be classified as overall high
18 risk of bias.¹⁷ Applying the recommended Cochrane criteria on Cipriani et al.’s ratings, there was
19 one trial at low risk of bias, 383 trials (73%) at unclear risk, and 138 trials (26%) at high risk of bias.
20 When we used our classifications for the blinding domains (i.e. all placebo-controlled trials were
21 rated as unclear risk of bias, and for the selective outcome reporting domain) there were no (0%)
22 trials at low risk of bias, 261 (50%) trials of unclear risk, and 261 (50%) trials of high risk of bias (S1
23 Appendix). If the three blinding domains were rated as high risk of bias in the placebo-controlled
24 trials, rather than unclear risk of bias, there were no (0%) trials at low risk, 108 trials (21%) of
25 unclear risk, and 414 trials (79%) of high risk of bias (S1 Appendix).
26
27
28
29

30 **Publication bias**

31 Publication bias of antidepressant trials is pervasive and distorts the evidence base.⁹ Many
32 industry funded antidepressant trials remain unpublished or are inadequately reported.⁹ Cipriani et
33 al.⁵ included 436 published and 86 unpublished studies, but as many as a thousand antidepressant
34 studies may have been conducted.¹³ We did a random effects meta-analysis of the placebo
35 comparisons according to publication status. The average effect size was lower in unpublished
36 studies (SMD 0.15 (95% CI: 0.11 to 0.19, 96 comparisons, 57 trials)) than in published studies
37 (SMD 0.33 (95% CI: 0.30 to 0.35, 294 comparisons, 196 trials) ($p < 0.0001$ for difference between
38 the two estimates). Our findings are very similar to those reported by Turner et al.⁹ in 2008 of
39 published versus unpublished antidepressant trials registered by the FDA who found an SMD of
40 0.37 (95% CI: 0.33 to 0.41) for published studies and 0.15 (95% CI: 0.08 to 0.22) for unpublished
41 studies. This indicates that the reported effect sizes by Cipriani et al.⁵ are highly inflated due to
42 publication bias. They correctly downgraded their confidence in the evidence due to the risk of
43 publication bias, but they should have estimated the impact of publication bias on their effect
44 estimate.
45
46
47
48

49 **Trial duration and long-term effects**

50 Cipriani et al.⁵ extracted outcome data as close to eight weeks follow-up as possible within an
51 interval of four to 12 weeks⁵ but did not provide a rationale for this decision.¹⁴ The common clinical
52 practice is to prescribe antidepressants for much longer periods. In the Netherlands, 43% of SSRI
53 users receive treatment for 15 months or more,²⁶ while 68% of those who use antidepressants in
54 the US take them for two years or more, and 25% take them for more than 10 years.³ Although the
55
56
57
58
59
60

short trial duration was acknowledged by the authors as a limitation, the lack of clinical relevance of such short follow-up should have been highlighted and the confidence in the evidence should have been downgraded one level in the GRADE domain of 'indirectness'. The authors should have extracted outcome data according to length of treatment and follow-up to assess any decrease in the treatment effect. According to the trial characteristics reported by Cipriani et al.⁵ 12 of the 304 placebo-controlled trials lasted more than 12 weeks. However, the authors⁵ listed misleading trial lengths because only four of these 12 trials contained a continuous double-blind, placebo-controlled phase of more than 12 weeks duration (S2 Appendix). The two placebo-controlled trials with the longest follow-up included only 81 participants at 36 weeks (S2 Appendix). A further consequence of a short follow-up period is an underestimation of serious and non-serious adverse events.²⁷

Placebo run-in and inclusion of already treated patients

The placebo run-in design is flawed and distorts the estimates of benefits and harms (Table 1A). Cipriani et al. did not provide a clear definition of a 'placebo run-in',¹⁴ but they characterised 260 (50%) of the 522 included trials as having a 'placebo run-in', 182 (35%) trials as unclear, and 80 (15 %) trials as having no 'placebo run-in'.⁵ In random effects meta-analyses of the placebo-controlled trials with and without a 'placebo run-in' we found that the effect sizes differed between the groups with an SMD of 0.31 (95% CI: 0.28 to 0.34, 221 comparisons, 142 trials) in trials with a 'placebo run-in', an SMD of 0.29 (95% CI: 0.25 to 0.33, 120 comparisons, 79 trials) where 'placebo run-in' was unclear, and an SMD of 0.22 (95% CI: 0.16 to 0.28, 46 comparisons, 30 trials) in trials without a 'placebo run-in' ($p=0.05$ for the difference between the three estimates). In a further subgroup analysis of unpublished trials without 'placebo run-in' the effect size was very small (SMD 0.08, 95% CI: -0.27 to 0.11, 8 comparisons, 5 trials). These data support our a priori assumption that the 'placebo run-in' design inflates the effect estimate by increasing the benefits in the antidepressant group and increased harms in the placebo group (Table 1A). The use of the 'placebo run-in' design and its implications were not discussed by Cipriani et al.⁵

Table 1. 'Placebo run-in', minimal clinical significant difference and response as outcome.

A. 'Placebo run-in' and the inclusion of already treated patients distort the benefit-harm balance

Cipriani et al.⁵ did not provide a definition of 'placebo run-in', but it usually involves that the participants, before the randomisation, receive placebo, typically for about a week after which non-adherent participants and those who responded well to the placebo (often called "placebo-responders") are excluded from the trial. Participants already in treatment with antidepressants, including the study drug, are virtually always allowed to enter the trial, and commonly all participants are tapered off ongoing antidepressant medication during the 'placebo run-in'. This study design may impact the effect estimates of placebo-controlled trials and the benefit/harm balance through several mechanisms that all favour the drug over placebo:

- Participants treated with the study drug, or a similar drug, prior to inclusion and subsequently randomised to the drug will most likely tolerate it and experience fewer harms

1
2
3
4 compared to a drug naïve population (**reduced harms in the drug group**).

- 5
6 - Participants treated with an antidepressant before the trial and subsequently randomised to
7 placebo might experience withdrawal symptoms that are often misinterpreted as signs of
8 worsening of the depression or as adverse events.⁴² Withdrawal symptoms typically occur
9 within a few days after discontinuation but there is great clinical variation⁴² (**reduced**
10 **benefits and increased harms in the placebo group**).
- 11
12
13 - Participants already treated with an antidepressant and subsequently randomised to the
14 study drug might experience withdrawal symptoms during the placebo run-in that are
15 alleviated by the study drug.⁴² It could be misinterpreted as an improvement of the
16 depression (**increased benefits in the drug group**).
- 17
18

19 **B. 'Response rates' lack clinical meaning**

20
21 The 'response rate' is usually defined as the number of participants in a randomised clinical trial
22 who achieve a reduction of more than 50% of the total score on a standardised observer-rated
23 scale for depression, such as the Hamilton depression rating scale or the Montgomery-Åsberg
24 rating scale. 'Non-response' does not necessarily imply that the participant's condition has not
25 improved, but simply that the improvement is rated to be less than the 50% reduction. The
26 difference might be as little as one point on the rating scale between a 'responder' and a 'non-
27 responder'. Thus, patients classified as 'non-responders' may actually have shown substantial
28 improvement. The difference in 'response rates' between antidepressants and placebo does
29 therefore not indicate the difference in the number of people who have improved, but only the
30 difference in the number of participants whose improvement exceeded an arbitrarily defined, but
31 clinically irrelevant, threshold. In addition, by only focusing on the number of participants crossing
32 the 50% reduction threshold the participants whose conditions deteriorate during the trial are
33 ignored. Therefore, it is more clinically meaningful to look at the average effect estimate of the drug
34 compared to placebo.

35
36
37

38 **C. Minimal clinically relevant difference**

39
40 Cipriani et al. reported an overall effect estimate measured as a standardised mean difference of
41 0.3 between antidepressants and placebo.⁵ The British National Institute of Health and Clinical
42 Excellence suggested in 2004 a difference of three points on the Hamilton depression rating scale,
43 or a standardised mean difference of 0.5, as a clinically significant change.⁶ However, this
44 difference was arbitrary and not based on empirical data.⁴³ Leucht et al. used clinical trial data in
45 2013 to suggest that clinicians are unable to detect reductions on the Hamilton depression rating
46 scale of three points or less.⁴⁴ Others have interpreted the same data and suggested that changes
47 of seven points or more on the Hamilton scale, corresponding to a standardised mean difference of
48 at least 0.875, are necessary for a clinician to detect a minimal clinical improvement.⁴⁵ We found
49 that the mean difference between antidepressants and placebo on the 17-item Hamilton scale,
50 based on Cipriani et al.'s data⁵ was 1.97 points on a scale from 0 to 52. This mean difference falls
51 below what is considered a clinically relevant effect.

52
53
54
55
56
57
58
59
60

Dropout as a proxy for harms

Overall dropout rates and dropouts due to adverse effects were assessed as measures of “acceptability” and “tolerability”, respectively, whereas the antidepressants’ actual harms, and serious and non-serious adverse events, were not assessed. It can be meaningful to use total dropout rates as a measure of the overall benefit/harm balance, but due to the biases introduced by including participants who are already known to tolerate an antidepressant drug and the use of ‘placebo run-in’, this outcome is biased in favour of the active drug (Table 1A). Furthermore, by omitting a careful analysis of the most serious harms, which include aggression, suicide and death,²⁸ and of specific adverse events, the review provided no basis for balancing benefits and harms, which is essential for informed consent and shared clinical decision-making and for evaluating whether the drugs have any clinical value. Adverse effects are common and a recent meta-analysis of 131 trials of SSRIs for depression found an increased risk of serious adverse events compared to placebo (OR 1.37; 95% CI: 1.08 to 1.75).⁷ This is likely an underestimate, as only 44 of the 131 included trials reported these data⁷ and as serious harms, including death, of antidepressants are often not reported in published papers.²⁹

Except for two drugs none of the included antidepressants had statistically significant lower total dropout rates than placebo,⁵ suggesting that the benefits do not outweigh the harms of the drugs compared to placebo. Cipriani et al.’s estimate is likely an underestimate since the total dropout rates were missing in 58 (11%) trials and the dropout rates due to adverse events were missing in 93 (18%) trials. A meta-analysis of dropouts in 73 trials based on clinical study reports obtained from drug regulators, rather than published data, showed that 12% more participants dropped out on antidepressants than on placebo.³⁰ This suggests that, seen from the patients’ perspective, placebo is a better drug than an antidepressant.

We identified clinical study reports for 19 of the 522 trials included in Cipriani et al.’s review. The outcomes of total dropout rates and dropout rates due to adverse events were fully reported in the clinical study reports. In comparison with those data, total dropout rates or dropouts due to adverse events were either not reported or incorrectly reported by Cipriani et al. in 12 (63%) of the 19 trials: total dropout rates were not reported for two trials and incorrectly reported for seven trials, dropouts due to adverse events were not reported for five trials and incorrectly reported for three trials (S1 Table). Our findings show that Cipriani et al.’s data were inaccurate and their estimates may therefore be incorrect because they relied on published data. They should have included clinical study reports because they are the most reliable source of trial data.

Lack of patient relevant outcomes

Patient relevant outcomes such as quality of life and sick leave are rarely measured and reported in psychiatric drug trials. Instead, the trials mostly rely on investigator-rated symptom scores, although validated self-rated symptom scales also exist. In a systematic review of SSRIs for depression in adults, only six of 131 trials reported quality of life data⁷ and even clinical study reports are unreliable because of selective reporting of this outcome.³⁰ The inability to cope with daily activities and the drugs’ side-effects may be more important to patients than their depressed mood³¹ and the decision not to specify patient-relevant outcomes in the protocol¹⁴ is a major limitation of the evidence and of Cipriani et al.’s overall conclusion.⁵

Clinically irrelevant efficacy outcomes

The network meta-analysis' primary efficacy outcome was 'response rate' (Table 1B). It is a problematic outcome because it lacks clinical relevance and it may create an illusion of clinical effectiveness.³² Dichotomisation of outcomes measured on rating scales leads to loss of statistical power, and it increases the risk of false positive results³³ and spuriously inflated effect sizes.³² Therefore, methodologists discourage the use of such dichotomised outcomes and they should generally be avoided when rating scale data are available.³³ These issues also apply to the review's secondary outcome of 'remission rates'. The choice made by Cipriani et al.⁵ to report only the relative odds ratios and not the trials' absolute 'response rates' has been criticised.³⁴ However, even absolute 'response rates' are of limited clinical relevance. Cipriani et al.⁵ did not address the problems related to 'response' and 'remission rates'.

Statistical versus clinical significance

Cipriani et al.⁵ reported the standardised mean difference (SMD) on a rating scale, which is more meaningful than the dichotomised outcomes.^{32 33} They reported an overall SMD for antidepressants versus placebo of 0.30 (95% credible interval: 0.26 to 0.34), but the number of trials and comparisons were unclear.⁵ We found a similar overall SMD for antidepressants versus placebo for the direct pair-wise comparisons of 0.29 (95% CI: 0.27 to 0.31, 390 comparisons, 253 studies). These effect estimates are statistically significant, but much below what is considered a clinically relevant effect (Table 1C). We calculated an overall mean difference for the trials that reported endpoint or change scores on the 17-item Hamilton depression rating scale (HAMD17), which was the most commonly used scale in the included trials (S2 Table). The mean difference between antidepressants and placebo was 1.97 points (95% CI: 1.74 to 2.21, 166 comparisons, 109 trials) on the HAMD17. This mean difference on the Hamilton scale also falls considerably below what is considered a clinically relevant effect (Table 1C). Cipriani et al. did not discuss the lack of clinical significance of their reported effect sizes.⁵

Selected, non-representative study populations

Antidepressant trials typically have extensive exclusion criteria that limit their external validity. These include psychiatric comorbidities, alcohol abuse, long duration of illness, and 'non-response' to previous antidepressant treatment.³⁵ The vast majority of patients in a clinical setting would not be eligible to enter randomised trials due to the exclusion criteria,³⁶ and the evidence coming from such trials is therefore of limited relevance. Furthermore, exclusion of previous 'non-responders' and inclusion of those who are expected to respond more favourably to treatment will be expected to bias the trials (Table 1A). None of these issues were considered by Cipriani et al.⁵ but should have resulted in downgrading of the confidence in the evidence in the GRADE domain of indirectness.¹⁸

The confidence in the evidence

Cipriani et al.⁵ assessed the certainty of evidence for the two main outcomes using the GRADE approach adapted for network meta-analyses. They provided the GRADE results for the head-to-head comparisons but we were unable to find the results for the placebo comparisons.⁵ Following the many issues related to the quality of the evidence, the certainty of evidence for the placebo comparisons should be downgraded two levels due to a 'high risk' of bias, two (or arguably) three levels in the domain of 'indirectness' due to short trial lengths, strict inclusion

1
2
3
4 criteria, and the use of 'placebo run-in', in addition to downgrading one level due to publication bias
5 as acknowledged by Cipriani et al.⁵ Downgrading due to the inherent indirectness of the network
6 meta-analysis should also be considered.³⁷ Taken together, the certainty of evidence can only be
7 'very low'.¹⁸
8

9 Discussion

10 We have identified a large number of important biases that were not taken into account in the
11 meta-analysis by Cipriani et al.⁵ The reported effect of antidepressants over placebo measured on
12 depression rating scales was inflated by several methodological flaws, and even if it is accepted it
13 is not clinically relevant. We show for the first time that the 'placebo run-in' study design works
14 toward producing inflated effect sizes, in addition to publication bias and other flaws. Furthermore,
15 the supporting evidence for long-term treatment is non-existing although most patients are treated
16 for years.¹²⁶ The certainty of evidence should be very low and our review has shown that the true
17 effect of antidepressants compared to placebo is likely much smaller than reported by Cipriani et
18 al.⁵ In fact, we cannot exclude the possibility that the average effect of antidepressants on
19 depression in adults might be negative. Taken together, Cipriani et al.'s conclusion⁵ that
20 antidepressants are more efficacious than placebo for depression is not supported by the
21 evidence.
22
23
24

25 Our review and meta-analyses relied on the data reported by Cipriani et al.⁵ It may be perceived as
26 a limitation that we did not perform our own separate systematic literature search and data
27 extraction. However, considering the multiple methodological flaws and the other problems we
28 have identified it would be necessary to analyse data based on the original clinical study reports to
29 make a more reliable assessment of the benefits and harms of antidepressants.
30
31

32 Previous meta-analyses (Figure 1) have found similar improvement in symptom scores as Cipriani
33 et al.⁵ Several of these reviews have considered carefully the methodological flaws, assessed the
34 harms, and have drawn different conclusions.⁶⁻⁸ We found that Cipriani et al. did not use standard
35 Cochrane methods for assessing the risk of bias, although stating to do so, and their results were
36 non-transparently presented.⁵ Most of the review's results cannot be reproduced because basic
37 information, such as the number of included studies, arms and participants for each meta-analysis,
38 was not reported.
39
40

41 The network meta-analysis methodology may hold some promise, but only in areas where clearly
42 effective interventions exist and need to be ranked, and the many statistical options should never
43 overshadow an initial critical assessment of the evidence and a clear presentation of the results. It
44 is misleading to rank the antidepressants when we have very low confidence that they are better
45 than placebo. Interestingly, our pairwise meta-analysis of improvement in symptom scores yielded
46 very similar results to those reported by Cipriani et al. The added benefit of the network meta-
47 analysis methodology is unclear.⁵
48
49

50 Our results highlight that the many hundreds of placebo-controlled trials of antidepressants have
51 failed to address and answer the most basic, patient-relevant questions regarding antidepressants'
52 benefits and harms. Although this has been known for years,¹³ it has not led to any changes in
53 research practice. Observational studies indicate that the effectiveness of antidepressants in
54 practice is very low. In the large, publicly funded, Sequenced Treatment Alternatives to Relieve
55
56
57

1
2
3
4 Depression (STAR*D) study, only 3% of the 4041 enrolled patients were considered “in remission”
5 after one year.³⁸
6

7 Misleading reviews such as the one by Cipriani et al.⁵ have the effect that they may prevent people
8 from seeking other solutions to alleviate their condition, such as psychotherapy and dealing with
9 psychosocial stressors and they may stall funding and research of such treatment modalities.
10

11 Our review has two implications: First, the review by Cipriani et al.⁵ and its conclusion should be
12 carefully revisited. In the light of our findings the review should not inform clinical practice. Second,
13 there is a need for a radical change in the way antidepressant trials are being conducted, reported,
14 and interpreted. Doctors, patients, peers, and politicians need to acknowledge the limitations of the
15 current evidence of antidepressants for depression and collectively act towards a better psychiatry.
16 To get reliable answers about the antidepressants’ benefits and harms in adults with depression,
17 we need large-scale, industry-independent, active placebo-controlled, long-term trials of drug naïve
18 participants, with patient relevant outcomes rather than ranking scales.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure legends

Figure 1. Previous meta-analyses reporting effect sizes for antidepressants versus placebo in adults.

Data are reported as standardised mean differences with 95% confidence intervals.

NICE 2004⁶: SSRIs. Kirsch 2008⁸: 'New generation' antidepressants. Turner 2008⁹: All antidepressants. Arroll 2009³⁹: Antidepressants for depression in primary care. Data represent a pooled estimate of tricyclic antidepressants and SSRIs versus placebo, fixed effects model. Fournier 2010⁴⁰: All antidepressants. Data represent pooled estimate from three groups of severity (mild to moderate, severe, very severe), fixed effects model. Gibbons 2012⁴¹: Fluoxetine and venlafaxine. Jakobsen 2017⁷: SSRIs. The effect size of mean change scores. Cipriani 2018⁵: All antidepressants.

For peer review only

Contributors

The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. All authors contributed to the conception and design of the study, advised on all statistical aspects, and interpreted the data. All authors analysed the data, drafted and reviewed the manuscript and approved the final version to be published. All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. KM is the guarantor.

Data sharing

All data files and the code for the statistical analyses are available from the Open Science Framework database (Access for peer review: https://osf.io/3prz9/?view_only=45a0afe9a09d4e9e9eb8a2c2d8d4bdc6. A DOI will be assigned after acceptance of the manuscript for publication).

Funding

The study was funded by the Nordic Cochrane Centre. The funder had no role in the design, data collection, analysis, interpretation of data, writing of the report, or in the decision to submit the article for publication. The research was designed, conducted, analysed, and interpreted by the authors entirely independently of the funding source.

Ethical approval

Not required.

Competing interests

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

Acknowledgments

We are grateful to Luis Carlos Saiz, Pharm D, PhD for valuable comments.

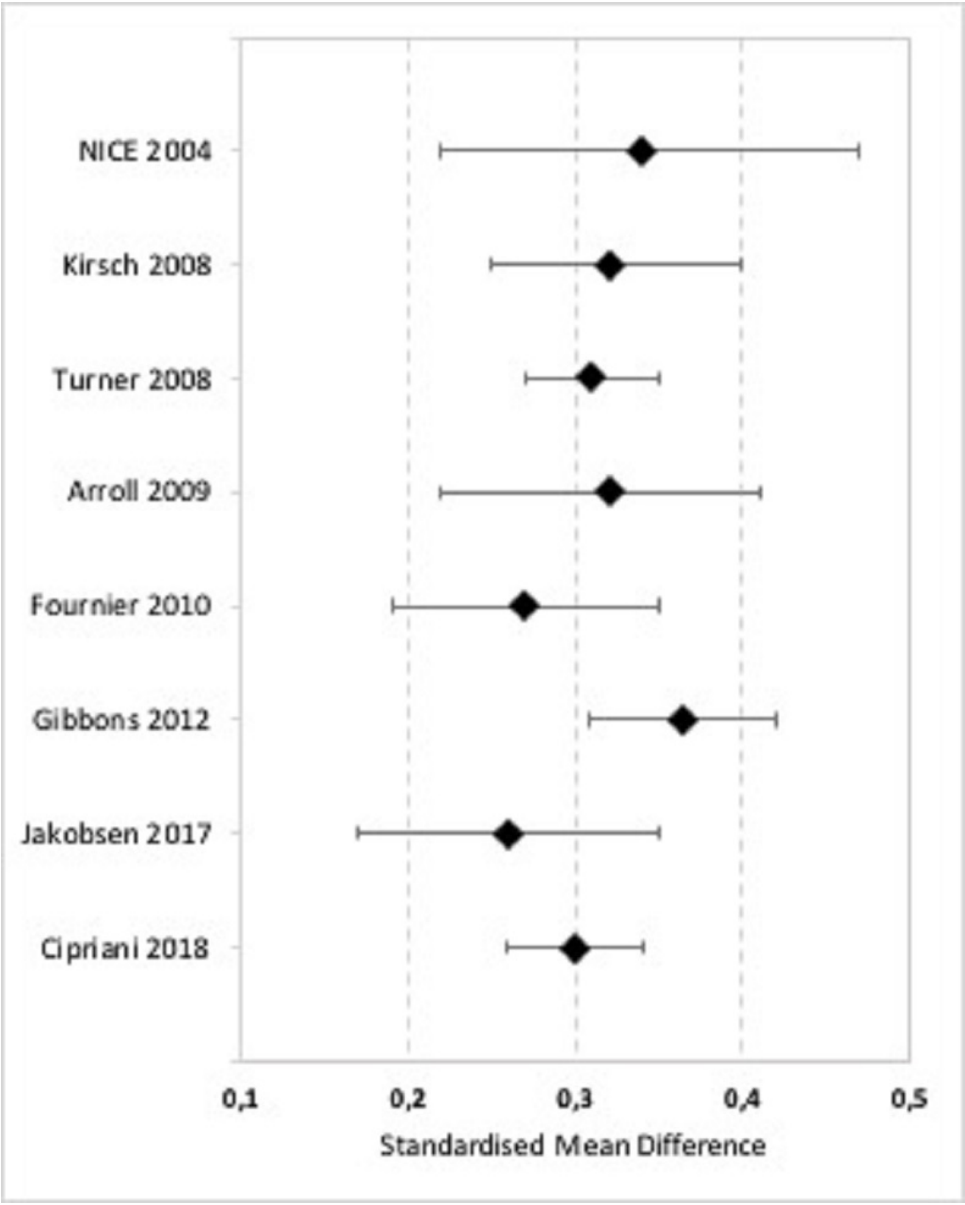
References

1. World Health Organisation. Media Centre. Depression: Fact Sheet. 2017 [Available from: <http://www.who.int/mediacentre/factsheets/fs369/en/> accessed 1 May 2018.
2. Danish Health Authority. MedicinForbrug - Indblik 2017: Laveste antal brugere af antidepressiv medicin de seneste 10 år [Medicine usage - 2017: Lowest number of antidepressant users in the last 10 years] 2017 [Available from: [https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/laegemidler/analyser-om-antidepressiva/laveste-antal-brugere-af-antidepressiv-medicin-de-seneste-10-aar-\(pdf\)-2017.pdf?la=da](https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/laegemidler/analyser-om-antidepressiva/laveste-antal-brugere-af-antidepressiv-medicin-de-seneste-10-aar-(pdf)-2017.pdf?la=da) accessed 1 May 2018.
3. U.S. Department of Health and Human Services: Centers for Disease Control and Prevention. Antidepressant Use Among Persons Aged 12 and Over: United States, 2011–2014 2017 [Available from: <https://www.cdc.gov/nchs/data/databriefs/db283.pdf> accessed 1 May 2018.
4. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord* 2018 doi: 10.1111/bdi.12609 [published Online First: 2018/03/15]
5. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
6. National Institute for Health and Care Excellence. Depression: Management of Depression in Primary and Secondary Care. Clinical Practice Guideline Number 23. London: National Institute for Health and Clinical Excellence; 2004 [Available from: <http://webarchive.nationalarchives.gov.uk/20080612001911/https://www.nice.org.uk/guidance/index.jsp?action=download&o=29617> accessed 1 May 2018.
7. Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC psychiatry* 2017;17(1):58. doi: 10.1186/s12888-016-1173-2 [published Online First: 2017/02/10]
8. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5(2):e45. doi: 10.1371/journal.pmed.0050045 [published Online First: 2008/02/29]
9. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358(3):252-60. doi: 10.1056/NEJMs065779 [published Online First: 2008/01/18]
10. The Guardian. The drugs do work: Antidepressants are effective, study shows: The Guardian; [updated 21 February 2018. Available from: <https://www.theguardian.com/science/2018/feb/21/the-drugs-do-work-antidepressants-are-effective-study-shows> accessed 1 May 2018.
11. TIME. These antidepressants are most effective, study says Time Magazine: Time Magazine; [updated 21 February 2018. Available from: <http://time.com/5169013/antidepressants-more-effective-placebo-treating-depression/> accessed 1 May 2018.
12. Hengartner MP. Methodological flaws, conflicts of interest, and scientific fallacies: Implications for the evaluation of antidepressants' efficacy and harm. *Front Psychiatry* 2017;8:275. doi: 10.3389/fpsy.2017.00275 [published Online First: 2017/12/23]
13. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14. doi: 10.1186/1747-5341-3-14 [published Online First: 2008/05/29]
14. 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(7):e010919. doi: 10.1136/bmjopen-2015-010919 [published Online First: 2016/07/13]

15. Gotzsche PC, Jorgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686. doi: 10.1136/bmj.d2686 [published Online First: 2011/05/12]
16. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC medical research methodology* 2014;14:25. doi: 10.1186/1471-2288-14-25 [published Online First: 2014/02/20]
17. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Available from <http://handbook-5-1.cochrane.org/> (accessed 1 May 2018): The Cochrane Collaboration, 2011.
18. Schünemann H, Brożek J, Guyatt G, et al. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. Available from guidelinedevelopment.org/handbook2013.
19. Page MJ, Higgins JP, Clayton G, et al. Empirical evidence of study design biases in randomized trials: Systematic review of meta-epidemiological studies. *PLoS One* 2016;11(7):e0159267. doi: 10.1371/journal.pone.0159267 [published Online First: 2016/07/12]
20. Perlis RH, Ostacher M, Fava M, et al. Assuring that double-blind is blind. *Am J Psychiatry* 2010;167(3):250-2. doi: 10.1176/appi.ajp.2009.09060820 [published Online First: 2010/03/03]
21. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* 2004(1):CD003012. doi: 10.1002/14651858.CD003012.pub2 [published Online First: 2004/02/20]
22. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291(20):2457-65. doi: 10.1001/jama.291.20.2457 [published Online First: 2004/05/27]
23. Woolley SB, Cardoni AA, Goethe JW. Last-observation-carried-forward imputation method in clinical efficacy trials: review of 352 antidepressant studies. *Pharmacotherapy* 2009;29(12):1408-16. doi: 10.1592/phco.29.12.1408 [published Online First: 2009/12/02]
24. Dziura JD, Post LA, Zhao Q, et al. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med* 2013;86(3):343-58. [published Online First: 2013/09/24]
25. Lundh A, Lexchin J, Mintzes B, et al. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017;2:MR000033. doi: 10.1002/14651858.MR000033.pub3 [published Online First: 2017/02/17]
26. Huijbregts KM, Hoogendoorn A, Slottje P, et al. Long-term and short-term antidepressant use in general practice: Data from a large cohort in the netherlands. *Psychother Psychosom* 2017;86(6):362-69. doi: 10.1159/000480456 [published Online First: 2017/11/14]
27. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012;13:138. doi: 10.1186/1745-6215-13-138 [published Online First: 2012/08/22]
28. Sharma T, Guski LS, Freund N, et al. Suicidality and aggression during antidepressant treatment: systematic review and meta-analyses based on clinical study reports. *BMJ* 2016;352:i65. doi: 10.1136/bmj.i65 [published Online First: 2016/01/29]
29. Hughes S, Cohen D, Jaggi R. Differences in reporting serious adverse events in industry sponsored clinical trial registries and journal articles on antidepressant and antipsychotic drugs: a cross-sectional study. *BMJ Open* 2014;4(7):e005535. doi: 10.1136/bmjopen-2014-005535 [published Online First: 2014/07/11]
30. Sharma T. Effects of selective serotonin reuptake inhibitors (SSRIs) and serotonin - norepinephrine reuptake inhibitors (SNRIs) on suicidality, violence, and quality of life (PhD Thesis). Available from: http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_ph_d.pdf. University of Copenhagen, 2018.
31. Eiring O, Landmark BF, Aas E, et al. What matters to patients? A systematic review of preferences for medication-associated outcomes in mental disorders. *BMJ Open*

- 2015;5(4):e007848. doi: 10.1136/bmjopen-2015-007848 [published Online First: 2015/04/10]
32. Kirsch I, Moncrieff J. Clinical trials and the response rate illusion. *Contemp Clin Trials* 2007;28(4):348-51. doi: 10.1016/j.cct.2006.10.012 [published Online First: 2006/12/22]
33. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;332(7549):1080. doi: 10.1136/bmj.332.7549.1080 [published Online First: 2006/05/06]
34. McCormack J, Korownyk C. Effectiveness of antidepressants. *BMJ* 2018;360:k1073. doi: 10.1136/bmj.k1073 [published Online First: 2018/03/11]
35. Posternak MA, Zimmerman M, Keitner GI, et al. A reevaluation of the exclusion criteria used in antidepressant efficacy trials. *Am J Psychiatry* 2002;159(2):191-200. doi: 10.1176/appi.ajp.159.2.191 [published Online First: 2002/02/02]
36. Zetin M, Hoepner CT. Relevance of exclusion criteria in antidepressant clinical trials: a replication study. *J Clin Psychopharmacol* 2007;27(3):295-301. doi: 10.1097/JCP.0b013e318058263f [published Online First: 2007/05/16]
37. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--indirectness. *J Clin Epidemiol* 2011;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014 [published Online First: 2011/08/02]
38. Pigott HE. The star*d trial: It is time to reexamine the clinical beliefs that guide the treatment of major depression. *Can J Psychiatry* 2015;60(1):9-13. doi: 10.1177/070674371506000104 [published Online First: 2015/04/18]
39. Arroll B, Elley CR, Fishman T, et al. Antidepressants versus placebo for depression in primary care. *Cochrane Database Syst Rev* 2009(3):CD007954. doi: 10.1002/14651858.CD007954 [published Online First: 2009/07/10]
40. 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. doi: 10.1001/jama.2009.1943 [published Online First: 2010/01/07]
41. Gibbons RD, Hur K, Brown CH, et al. Benefits from antidepressants: synthesis of 6-week patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and venlafaxine. *Arch Gen Psychiatry* 2012;69(6):572-9. doi: 10.1001/archgenpsychiatry.2011.2044 [published Online First: 2012/03/07]
42. Fava GA, Gatti A, Belaise C, et al. Withdrawal symptoms after selective serotonin reuptake inhibitor discontinuation: A systematic review. *Psychother Psychosom* 2015;84(2):72-81. doi: 10.1159/000370338 [published Online First: 2015/02/28]
43. Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336(7643):516-7. doi: 10.1136/bmj.39510.531597.80 [published Online First: 2008/03/06]
44. Leucht S, Fennema H, Engel R, et al. What does the HAMD mean? *J Affect Disord* 2013;148(2-3):243-8. doi: 10.1016/j.jad.2012.12.001 [published Online First: 2013/01/30]
45. Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of antidepressant-placebo differences. *Contemp Clin Trials* 2015;43:60-2. doi: 10.1016/j.cct.2015.05.005 [published Online First: 2015/05/17]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



66x83mm (300 x 300 DPI)

1
2
3
4
5
6
7
8
9

Flaws in the evidence base of antidepressants for depression: A reanalysis and critical review of a network meta-analysis

10 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

11
12 ^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
13
14
15

16
17 S1 Table
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

S1 Table. Comparison between the outcomes of total dropout rates and dropouts due to adverse events as reported by Cipriani et al. [1] and the data reported in the clinical study reports.

Trial ID (From Cipriani et al. dataset)	Total dropouts reported by Cipriani et al. compared with clinical study reports*	Dropouts due to adverse events reported by Cipriani et al. compared with clinical study reports**
Goldstein2002 (HMAQ - Study Group A)	Match	Match
Study F1J-MC-HMAQ - Study Group B (starting page in the pdf document: 147)	Match	Not reported
Goldstein2004a (HMAQ - Study Group A, ID#4091)	Discrepancy	Match
Goldstein2004b (HMAQ - Study Group B, ID#4091)	Match	Discrepancy
Detke2004 (HMAQ Study Group A)	Match	Match
Perahia2006 (HMAQ - Study Group B)	Match	Match
Detke2002a (HMBH - Study Group A)	Match	Match
Detke2002b (HMBH - Study Group B)	Match	Discrepancy
0600B-367	Not reported	Not reported
0600B1-384	Discrepancy	Discrepancy
VEN 600A-303 (FDA)	Discrepancy	Match
Cunningham1994 (VEN 600A-302 FDA)	Discrepancy	Match
VEN 600A-313 (FDA)	Discrepancy	Match
0600A1-372	Not reported	Not reported
Thase1997 (VEN XR 209 FDA)	Match	Match
Rudolph1998 (VEN 600A-203 (FDA)	Match	Match
SER 101 (FDA)	Match	Not reported
Fabre1995 (SER 103 FDA)	Match	Match
SER 310 (FDA)	Discrepancy	Not reported

*Our comparison considered all treatment arms in the trials. If data for all arms matched, we assigned an assessment of 'match'; if data for any arm differed but were reported, we assigned an assessment of 'discrepancy'; if data for any arm was not reported, we assigned an assessment of 'not reported'.

References

1. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, 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. Epub 2018/02/27. doi: 10.1016/S0140-6736(17)32802-7. PubMed PMID: 29477251.

For peer review only

1
2
3
4
5 **Flaws in the evidence base of antidepressants for depression: A**
6 **reanalysis and critical review of a network meta-analysis**
7
8
9

10 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

11
12 ^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
13
14
15

16
17 **S2 Table**
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

S2 Table. Rating scales used in the included placebo-controlled trials (n = 304) [1].

Scale	Number of trials	Number of treatment arms
HAMD17	137	198
HAMD21	92	139
HAMD24	13	22
HAMD29	1	2
HAMD31	1	2
HAMD unspecified	22	25
IDS-IVR-30	1	1
MADRS	29	50
NA	8	12

NA: not available; HAMD17: 17-item Hamilton depression rating scale; IDS-IVR-30: 30-item Inventory of Depressive Symptomatology-Self Report; MADRS: Montgomery-Åsberg Depression Rating Scale.

For peer review only

References

1. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, 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. Epub 2018/02/27. doi: 10.1016/S0140-6736(17)32802-7. PubMed PMID: 29477251.

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Flaws in the evidence base of antidepressants for depression: A reanalysis and critical review of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Appendix

For peer review only

Risk of bias

Blinding of the participants

Cipriani et al. [1] classified three [2-4] out of the 522 trials as being at low risk of bias in the “blinding of participants” domain:

Brunoni et al. [2] tested the blinding, by asking the participants who completed the trial to guess their allocation. They reported that 39 (75%) of 52 participants on placebo, and 29 (58%) of 50 participants on sertraline were able to correctly guess their allocation. Brunoni et al. suggested that the results were “driven by clinical improvement... rather than blinding failure” [2]. We disagree and would categorise this trial at high risk of blinding bias of the participants.

Edwards et al. [3] did not describe how and when they tested the blinding, and two of the included participants were not asked. They reported that 12 (60%) of 20 participants on placebo and 12 (63%) of 19 participants on paroxetine were able to correctly guess their treatment allocation. Edwards et al. concluded that their results “confirmed the blindness of the study”, but we would categorise the trial to be at unclear risk of bias.

Schatzberg et al. [4] did not test the blinding, and it is unclear why this trial was rated at low risk of bias, rather than the “stated but not tested” categorisation.

Summary risk of bias assessments

Criteria for assessments

To categorise the 522 trials included by Cipriani et al. [1], we followed the Cochrane Handbook’s criteria for an overall risk of bias assessment [5]. Each domain in the risk of bias tool likely affects all five included outcomes assessed by Cipriani et al., and we therefore considered all bias domains as “key domains”, according to table 1.

Table 1. Criteria for overall risk of bias assessment.

Low risk of bias	All key domains classified as low risk of bias
Unclear risk of bias	One or more key domains classified as unclear risk of bias, and no domains classified as high risk of bias.
High risk of bias	One or more key domains classified as high risk of bias

We collapsed the three blinding domains in our Excel dataset and used the following criteria for our categorisation: All placebo-controlled trials were classified as unclear risk of bias in the main analysis and as high risk of bias in the sensitivity analysis. Trials that only contained head-to-head antidepressant arms, and no placebo arm, were rated as low risk of bias if the three blinding domains were rated as ‘low’ or ‘stated but not tested’ by Cipriani et al. [1]. We rated the collapsed blinding domain as unclear risk of bias, if one or more of the blinding domains were rated as unclear by Cipriani et al. Trials with missing data for any of the five included outcomes were categorised as high risk of bias. For the remaining bias domains, we adopted the categorisations by Cipriani et al. [1]. Our results are compared with Cipriani et al.’s in table 2.

Table 2. Comparison of the overall risk of bias assessments.

Cipriani et al. overall assessment		Our overall assessments			
		Cochrane categorisation	Sensitivity analysis using the Cipriani et al. categorisation	Our assessment	Sensitivity analysis of the blinding domains
Low risk	46 trials (9%)	Low	1 trial (0.2%)	0 trials (0%)	0 trials (0%)
“Moderate” risk	380 trials (73%)	Unclear	383 trials (73%)	261 trials (50%)	108 (21%)
High risk	96 trials (18%)	High	138 trials (26%)	261 trials (50%)	414 (79%)

References

1. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, 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. Epub 2018/02/27. doi: 10.1016/S0140-6736(17)32802-7. PubMed PMID: 29477251.
2. Brunoni AR, Valiengo L, Baccaro A, Zanao TA, de Oliveira JF, Goulart A, et al. The sertraline vs. electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial. *JAMA psychiatry*. 2013;70(4):383-91. Epub 2013/02/08. doi: 10.1001/2013.jamapsychiatry.32. PubMed PMID: 23389323.
3. Edwards JG, Goldie A. Placebo-Controlled trial of paroxetine in depressive illness. *Human Psychopharmacology: Clinical and Experimental*. 1993;8(3):203-9. doi: 10.1002/hup.470080306.
4. Schatzberg AF, Kremer C, Rodrigues HE, Murphy GM, Jr., Mirtazapine vs. Paroxetine Study G. Double-blind, randomized comparison of mirtazapine and paroxetine in elderly depressed patients. *Am J Geriatr Psychiatry*. 2002;10(5):541-50. Epub 2002/09/06. PubMed PMID: 12213688.
5. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 [updated March 2011]. Available from <http://handbook-5-1.cochrane.org/> (accessed 1 May 2018): The Cochrane Collaboration; 2011.

1
2
3
4
5 **Flaws in the evidence base of antidepressants for depression: A**
6 **reanalysis and critical review of a network meta-analysis**
7
8
9

10 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

11
12 ^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
13
14
15

16 **S2 Appendix**
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Trial duration and long-term effect estimates

Overall trial length

According to the trial characteristics by Cipriani et al. 492 (94%) of the 522 included trials lasted between four and 12 weeks (figure 1), 28 trials (5%) lasted more than 12 weeks, and the trial duration was unclear for two trials (table 3).

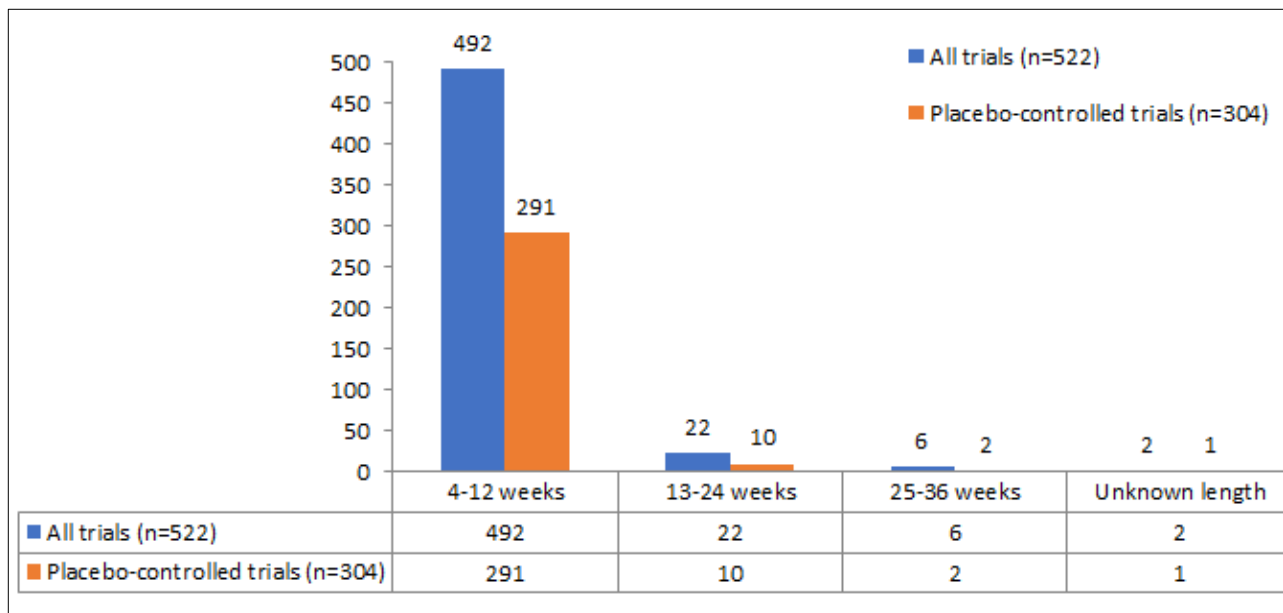


Figure 1. All 522 trials arranged into three follow-up periods according to Cipriani et al.'s trial characteristics: Four to 12 weeks (the period used by Cipriani et al.), 13 to 24 weeks, and 25 to 36 weeks [1].

Longer-term placebo-controlled trials

Of the 28 trials lasting more than 12 weeks, 12 trials had a placebo-controlled arm (table 3) and 16 were head-to-head trials. However, upon closer examination, eight of the 12 placebo-controlled trials consisted of several 'phases' and only four trials contained a continuous randomised placebo-controlled phase of more than 12 weeks (table 4). It is unclear why Cipriani et al. [1] did not list the correct trial length characteristics, since the various extension phases were clearly described in the available documents, also for the unpublished trials.

Table 3. All 522 trials arranged according to the trial length characteristics by Cipriani et al. [1].

Trial length according to Cipriani et al. [1] (N = 522)	Trials without a placebo arm (N)	Placebo-controlled trials (N)
4 weeks	9	39
5 weeks	11	4
6 weeks	92	100
7 weeks	9	0
8 weeks	59	118
9 weeks	3	2
10 weeks	4	10
12 weeks	14	18
13 weeks	1	1
16 weeks	2	2
24 weeks	0	1

25 weeks	9	6
26 weeks	3	0
36 weeks	1	0
Trial length unknown	0	2
Total	218	304

Table 4. Longer-term placebo-controlled trials (trial length of more than 12 weeks).

Trial	Drug arm 1	Drug arm 2	Trial length according to Cipriani et al. (weeks)	Actual trial length (weeks)	Study design and phases
Oakes 2012a [2]	duloxetine	-	36	36	-
Oakes 2012b [2]	duloxetine	-	36	36	-
CL3-20098-022 [3]	agomelatine	fluoxetine	24	6 + 18	"Optional double-blind placebo-controlled extension" (page 24) for "responders to treatment" (page 26)
CL3-20098-023 [3]	agomelatine	paroxetine	24	6 + 18	"Optional double-blind placebo-controlled extension" (page 24) for "W6 responders" (page 26)
CL3-20098-024 [3]	agomelatine	fluoxetine	24	6 + 18	"Optional double-blind placebo-controlled extension" (page 24) for "responders to treatment at W6" (page 26)
CL3-20098-026 [3]	agomelatine	-	24	6 + 18	"18-week extension period" of "responders at W6" (page 42)
CL3-20098-070 [4]	agomelatine	-	24	8 + 16	16-week extension period for responders
Robinson 2014 [5]	duloxetine	-	24	12 + 12	From week 12 to 20 placebo rescue or dose increase were available
Lopez-Rodriguez 2004 [6]	fluoxetine	-	20	20	-
Barber 2011 [7]	sertraline	-	16	8 + 8	Non-responders on sertraline were switched to venlafaxin ER at week 8
Dimidjian 2006 [8]	paroxetine	-	16	8 + 8	The blind was broken at week 8 and the placebo arm was offered other treatments
Leclubier 1997 [9]	venlafaxine	-	13	13	-

References

1. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, 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. Epub 2018/02/27. doi: 10.1016/S0140-6736(17)32802-7. PubMed PMID: 29477251.
2. Oakes TM, Myers AL, Marangell LB, Ahl J, Prakash A, Thase ME, et al. Assessment of depressive symptoms and functional outcomes in patients with major depressive disorder treated with duloxetine versus placebo: primary outcomes from two trials conducted under the same protocol. *Hum Psychopharmacol*. 2012;27(1):47-56. Epub 2012/01/14. doi: 10.1002/hup.1262. PubMed PMID: 22241678.
3. CHMP assessment report for Thymanax. EMEA/H/C/000916. 2008 [cited 2018 1 May]. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/000916/WC500038315.pdf.
4. EudraCT Number 2009-011795-29. 2009 [cited 2018 1 May]. Available from: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2009-011795-29/FI>.
5. Robinson M, Oakes TM, Raskin J, Liu P, Shoemaker S, Nelson JC. Acute and long-term treatment of late-life major depressive disorder: duloxetine versus placebo. *Am J Geriatr Psychiatry*. 2014;22(1):34-45. Epub 2013/12/10. doi: 10.1016/j.jagp.2013.01.019. PubMed PMID: 24314888.
6. López RJ, López BMA, Vargas TBE, Villamil SV. Estudio doble ciego con antidepressivo, psicoterapia breve y placebo en pacientes con depresión leve a moderada. *Salud Mental*. 2004;27(5):53-61.
7. Barber JP, Barrett MS, Gallop R, Rynn MA, Rickels K. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *J Clin Psychiatry*. 2012;73(1):66-73. Epub 2011/12/14. doi: 10.4088/JCP.11m06831. PubMed PMID: 22152401.
8. Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*. 2006;74(4):658-70. Epub 2006/08/03. doi: 10.1037/0022-006X.74.4.658. PubMed PMID: 16881773.
9. Lecrubier Y, Bourin M, Moon CA, Schifano F, Blanchard C, Danjou P, et al. Efficacy of venlafaxine in depressive illness in general practice. *Acta Psychiatr Scand*. 1997;95(6):485-93. Epub 1997/06/01. PubMed PMID: 9242843.

BMJ Open

Considering the flaws in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-024886.R1
Article Type:	Research
Date Submitted by the Author:	08-Sep-2018
Complete List of Authors:	Munkholm, Klaus; Nordic Cochrane Centre, Righospitalet, 7811 Paludan-Müller, Asger; Nordic Cochrane Centre, Righospitalet, 7811 Boesen, Kim; Nordic Cochrane Centre, Righospitalet, 7811
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Pharmacology and therapeutics, Evidence based practice
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

1
2
3
4
5 **Considering the flaws in the evidence base of antidepressants**
6 **for depression: A reanalysis of a network meta-analysis**
7
8
9
10

11 **Authors**

12 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a
13
14

15 **Author affiliations**

16 ^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43

44 **Correspondence to:**

45 Klaus Munkholm, M.D., DMSc.
46 Nordic Cochrane Centre
47 Rigshospitalet, 7811
48 Blegdamsvej 9
49 200 København Ø
50 Denmark
51 Phone: +45 35457112
52 E-mail: km@cochrane.dk
53
54
55
56
57
58
59
60

Abstract

Objectives

To investigate whether the conclusion of a recent systematic review and network meta-analysis (Cipriani et al. 2018) that antidepressants are more efficacious than placebo for adult depression was supported by the evidence.

Design

Reanalysis of a systematic review, with meta-analyses.

Data sources

522 trials (116 477 participants) as reported in the systematic review by Cipriani et al. (2018) and clinical study reports for 19 of these trials.

Analysis

We used the Cochrane Handbook's risk of bias tool and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to evaluate the risk of bias and the certainty of evidence, respectively. The impact of several study characteristics and publication status was estimated using pair-wise subgroup meta-analyses.

Results

Several flaws in the evidence base of antidepressants were either unrecognised or underestimated in the systematic review by Cipriani et al (2018). The effect size for antidepressants versus placebo on investigator-rated depression symptom scales was higher in trials with a 'placebo run-in' study design compared to trials without a 'placebo run-in' design ($p = 0.05$). The effect size of antidepressants was higher in published trials compared to unpublished trials ($p < 0.0001$). The outcome data reported by Cipriani et al. differed from the clinical study reports in 12 (63%) of 19 trials. The certainty of the evidence for the placebo-controlled comparisons should be very low according to GRADE due to a high risk of bias, indirectness of the evidence, and publication bias. The mean difference between antidepressants and placebo on the 17-item Hamilton depression rating scale (range 0 to 52 points) was 1.97 points (95% CI: 1.74 to 2.21).

Conclusions

The evidence does not support definitive conclusions regarding the benefits of antidepressants for depression in adults. It is unclear whether antidepressants are more efficacious than placebo.

Strengths and limitations of this study

- Empirical evidence was provided showing how many flaws in the evidence base for antidepressants for depression affect the apparent effect size for antidepressants
- For the first time, the impact of the 'placebo run-in' study design on the apparent effect size for antidepressants compared with placebo was estimated
- We reported the effect estimate of antidepressants compared with placebo as a mean difference on the investigator-rated Hamilton depression rating scale to provide an outcome measure that can be easily interpreted by patients and clinicians
- When possible, we compared the data reported by Cipriani et al. (2018) on the outcomes of total dropouts and dropouts due to adverse events with the clinical study reports that we have previously obtained from the European Medicines Agency
- Our analyses relied on the data reported in the systematic review by Cipriani et al. (2018) and we did not perform a separate literature search and data extraction; given the flaws we have identified, a reliable assessment would need to be based on clinical study reports and individual patient data

Introduction

WHO estimates that 300 million people globally suffer from depression, making depression the leading cause of disability worldwide.¹ In Denmark, 10% of all adults 25 years and older were in treatment with antidepressants in 2016.² In the US, 13% of persons 12 years and older were in treatment in 2014, making antidepressants one of the three most commonly used drug classes.¹ Prescriptions for antidepressants cost the National Health Service in the United Kingdom an estimated £267 million in 2016.³ Research that guides clinical treatment of depression therefore has a potentially important impact on millions of people and on national economies.

The recent network meta-analysis of antidepressants for depression by Cipriani et al.⁴ is the largest meta-analysis of antidepressants to date in terms of included studies and participants. It specifically aimed to inform clinical guidelines, patients, physicians and policy makers by comparing 21 antidepressants for the treatment of adults with depression. The review's primary outcomes were 'response rate' (defined as the number of participants with at least a 50% reduction on an observer-rated depression scale) and overall dropout rates. The secondary outcomes were depression symptom scores, 'remission rate' (defined as the number of participants with an observer-rated depression score below a certain threshold), and dropouts due to adverse events. Cipriani et al. found that all 21 antidepressants were more effective than placebo, whereas only two of the drugs had fewer dropouts compared with placebo. Based on these findings, they⁴ ranked the antidepressants according to 'response rate' and overall dropout rate and concluded that antidepressants were more efficacious than placebo in adults with major depressive disorder. The improvement in symptom scores they found were very similar to previous meta-analyses (Figure 1), some of which have concluded that the benefit of antidepressants is doubtful.⁵⁻⁸ The review received widespread media coverage, largely citing it as finally putting to rest any doubts regarding the efficacy of antidepressants,^{9,10} and the message of antidepressants being effective was strongly conveyed by some of the authors in the press,⁹ adding that the benefits outweigh side effects.¹⁰

There are many methodological issues and flaws in trials of antidepressant agents,¹¹ of which many have been acknowledged for decades.¹² Research aiming to inform clinical practice on the use of antidepressants for depression must recognise these limitations. We have already addressed some of the limitations in the risk of bias assessment in the Cipriani et al. review.¹³ However, given the potential implications of Cipriani et al.'s review,⁴ we here aimed to provide a more comprehensive assessment. Specifically, we wished to investigate how the flaws in the evidence base were addressed, whether the review's assessment of the risk of bias within the included trials and the evaluation of the certainty of evidence were appropriate and followed the authors' stated methods, and whether the conclusion was supported by the evidence. We furthermore aimed to provide empirical evidence on the impact of these methodological flaws by using the data reported by Cipriani et al.⁴

Methods

Data collection

We extracted the review's risk of bias assessments and descriptive data from the online supplement and converted the data to Microsoft Excel format. We downloaded the online dataset⁴ and merged the files for our statistical analyses.

1
2
3
4 We cross-referenced the included trials with the clinical study reports that we previously obtained
5 from the European Medicines Agency (EMA) in 2010.¹⁴ We compared the outcomes of total
6 dropout rates and dropouts due to adverse events as reported in the clinical study reports with the
7 data reported by Cipriani et al.⁴
8

9 **Statistical analyses**

10 Descriptive analyses were made in Microsoft Excel. We used the statistical software R (version
11 3.4.3) for random effects meta-analyses based on the inverse variance method and calculated
12 effect sizes as standardised mean differences (SMD) as Hedges' *g* with corresponding 95%
13 confidence intervals (95% CI). The extent of variation among the intervention effects observed in
14 different studies was calculated as Tau^2 and the percentage of the variability in effect estimates
15 that was due to heterogeneity was calculated as I^2 . For the comparisons between antidepressants
16 and placebo on rating scales, we used the Hartung-Knapp-Sidik-Jonkman approach because it
17 results in fewer type I errors than the DerSimonian and Laird approach.¹⁵ We based our analyses
18 on the number of participants from Cipriani et al.'s 'efficacy' analyses.⁴ In studies with more than
19 one drug arm the total number of participants in the placebo group was split evenly between the
20 active comparisons and the means and standard deviations were unchanged.¹⁶ We did subgroup
21 analyses based on the use of a 'placebo run-in' study design, sponsorship and publication status,
22 according to the trial characteristics published by Cipriani et al.⁴
23
24
25

26 **Quality assessments**

27 We evaluated whether Cipriani et al.'s risk of bias assessments were in accordance with the
28 Cochrane Handbook,¹⁶ as stated by the authors.⁴ Where the approach differed we compared the
29 risk of bias assessment by Cipriani et al.⁴ with our reassessment following the Cochrane
30 Handbook.¹⁶ The specific domains (and type of bias) assessed were sequence generation
31 (selection bias), allocation sequence concealment (selection bias), blinding of participants and
32 personnel (performance bias), blinding of outcome assessment (detection bias), incomplete
33 outcome data (attrition bias), selective outcome reporting (reporting bias) and other potential
34 sources of bias.¹⁶
35

36 We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE)¹⁷
37 approach to evaluate the certainty of evidence, which, for systematic reviews, reflects the extent of
38 confidence that an estimate effect is correct. GRADE considers five domains that affect the quality
39 of the evidence: the included trials' internal risk of bias, inconsistency of the included trials' results
40 and large heterogeneity, indirectness of the evidence due to poor external validity, imprecision of
41 the effect estimate and wide confidence intervals, and publication bias.¹⁷
42
43

44 **Patient involvement**

45 No patients were involved in the development of the research question, design and implementation
46 of the study, or interpretation of the results.
47
48
49

50 **Results**

51 **Risk of bias**

52 ***Randomisation sequence generation and allocation sequence concealment***

53
54
55
56
57
58
59
60

1
2
3
4 Cipriani et al.⁴ judged 426 (82%) and 460 (88%) of the 522 included trials to be of unclear risk of
5 bias with respect to randomisation sequence generation and allocation concealment, respectively.
6 The remaining trials were of low risk of bias. Trials at high or unclear risk of bias within these two
7 domains are likely to report inflated effect estimates, especially of subjective outcomes.¹⁸ Cipriani
8 et al. did not describe how they assessed the risk of bias in relation to the randomisation sequence
9 generation or the allocation concealment, and we were therefore unable to evaluate if their
10 methods followed those outlined in the Cochrane Handbook.¹⁶
11

Blinding of participants, personnel and outcome assessment

12
13 Cipriani et al.⁴ did not use the standard Cochrane categorisation of low, unclear or high risk of bias
14 due to a lack of blinding.¹⁶ They categorised instead 513 (98%) studies as 'stated-not tested' in at
15 least one of the three blinding domains, meaning that the trial had stated to be double-blind, but did
16 not test the blinding integrity. While this implied the presence of a blinding issue, their
17 categorisation did not affect the overall risk of bias assessment⁴ and it seemed that the 'stated-not
18 tested' domains were counted as 'low risk of bias'. Two of the three trials categorised by Cipriani et
19 al.⁴ to be at low risk of bias in the blinding of participants' domain had tested the blinding integrity
20 (S1 Appendix). The blinding was likely compromised in both trials. Adverse effects of
21 antidepressants are common and often reveal who receives active medication and who receives
22 placebo in a randomised trial. The degree of unblinding is extensive and leads to inflated effect
23 estimates,¹⁹ and smaller effects have been observed when the trials were better blinded by adding
24 atropine to the placebo.²⁰ Given these issues, all placebo-controlled trials should arguably be
25 categorised as at least unclear, or perhaps even at high risk of bias.
26
27
28

Incomplete outcome data

29
30
31 Cipriani et al. rated trials that used an appropriate imputation method as low risk of bias.⁴ Trials
32 that used an 'inappropriate' imputation method were rated according to several arbitrary cut-offs:
33 When the dropout rates were unbalanced between the arms, defined as more than a 5% difference
34 for the head-to-head comparisons and a 10% difference for the placebo comparisons, they were
35 rated as high risk of bias. When the dropout rates between the arms were not unbalanced but the
36 total dropout rate was more than 20% they were rated unclear, and if the total dropout rate was
37 less than 20% they were rated as low risk of bias. This method is not in accordance with the
38 Cochrane Handbook, which emphasises that it is not possible to formulate a simple rule for judging
39 a study to be at low or high risk of attrition bias in that the risk of bias depends on several factors.¹⁶
40 Further, the authors did not consider the reasons for dropout, although this is also recommended
41 by the Cochrane Handbook.¹⁶
42
43

44
45 According to Cipriani et al. 121 (23%) trials were at high risk of attrition bias, but we could not
46 replicate these results. The overall attrition rate was more than 20% in 334 (64%) trials. Using the
47 cut-offs defined by Cipriani et al. we found that the dropout rates were unbalanced between the
48 arms in 202 trials (39%) and according to the methods described by Cipriani et al.⁴ they should
49 have been rated as high risk of bias unless an "appropriate imputation method" was used. Cipriani
50 et al. characterised the last observation carried forward (LOCF) method as inappropriate,²¹ but
51 they did not provide data on the used imputation method in the included trials. We were therefore
52 not able to apply Cipriani et al.'s categorisations in our reassessment of the attrition bias. Most
53 antidepressant trials use the LOCF imputation method,²² which may lead to an underestimation of
54 the variability, a falsely low p-value, and an overestimation of treatment effects.²³
55
56
57
58
59
60

Selective outcome reporting

Cipriani et al.⁴ judged that 402 (77%) of the 522 trials were of low risk of outcome reporting bias, 100 (19%) of unclear risk, and 20 (4%) of high risk of bias. Their assessments were based on the reporting of the review's two primary outcomes of 'response rates' and overall dropout rates and a trial was only rated at high risk of bias in case both outcomes were missing. This is not in accordance with the Cochrane Handbook, in which a study-level judgement across all relevant outcomes is recommended.¹⁶ According to our analyses the review's three secondary outcomes of dropouts due to adverse events, depression symptoms measured on depression symptom scales, and 'remission rates' were not reported in 93 (18 %) trials, 98 (19%) trials, and 71 (14 %) trials, respectively. We found that a total of 182 (35%) trials did not report at least one primary or secondary outcome and, following the recommendation by the Cochrane Handbook to consider all relevant outcomes, these trials should probably have been rated as high risk of bias.¹⁶ Selective outcome reporting leads to overestimation of the benefits and underestimation of the harms of interventions.²⁴

Other bias domain

The authors omitted the 'other bias' domain from the risk of bias assessment although it is an integrated part of Cochrane's risk of bias tool.¹⁶ Relevant biases included in this domain were baseline imbalances and design-specific risks of bias for cross-over and cluster randomised trials, which were eligible according to the Cipriani et al. protocol,²¹ although the trial designs were not specified in the review.⁴ Some argue that 'vested interests' should also be considered, since industry sponsored drug studies lead to more favourable effects than other studies by mechanisms that are not explained by the usual bias domains.²⁵ We explored whether industry sponsorship was associated with larger effect estimates, by performing random effects meta-analyses of the placebo-controlled trials according to sponsorship using the categorisation by Cipriani et al. (S1 Appendix). We found a lower effect size in trials categorised as "sponsored" (SMD of 0.27 (95% CI: 0.25 to 0.30, 341 comparisons, 207 trials)) than in trials categorised as "unclear" (SMD of 0.39 (95% CI: 0.25 to 0.52, 12 comparisons, 10 trials)) and "not sponsored" (SMD of 0.41 (95% CI: 0.31 to 0.52, 37 comparisons, 36 trials) ($p=0.005$ for the difference between the three estimates) (Table 1).

Table 1. Random effects pairwise meta-analyses of antidepressants versus placebo.

	N trials	N comparisons	ES	95% CI	Tau ²	I ²
Overall (SMD)						
Overall	253	390	0.29*	0.27-0.31	0.038	40.1%
Overall (Mean difference on the HAMD17)						
Overall	109	166	1.97**	1.74-2.21	1.896	27.6%
Publication status						
Published	196	294	0.33*	0.30-0.35	0.037	40.0%
Unpublished	57	96	0.15*	0.11-0.19	0.020	0.0%
'Placebo run-in'						
Yes	142	221	0.31*	0.28-0.34	0.043	35.0%
Unclear	79	120	0.29*	0.25-0.33	0.032	47.6%

No	30	46	0.22*	0.16-0.29	0.032	35.5%
Sponsorship						
Sponsored	207	341	0.27*	0.25-0.30	0.033	35.4%
Unclear	10	12	0.39*	0.25-0.52	0.026	33.0%
Not sponsored	36	37	0.41*	0.31-0.52	0.075	55.7%

ES: effect size; *: standardised mean difference (SMD); **: mean difference on 17-item Hamilton depression rating scale; HAMD17: 17-item Hamilton depression rating scale; 95% CI: 95% confidence interval; Tau²: estimate of overall heterogeneity; I²: inconsistency.

Summary risk of bias assessment

The authors deviated from Cochrane's overall risk of bias categorisation of low, unclear or high risk of bias,¹⁶ by introducing their own category of 'moderate' risk of bias. They classified the trials as low risk of bias if none of the domains assessed were rated as high risk of bias and three or less were rated as unclear risk; moderate if one domain was rated as high risk of bias or none were rated as high risk of bias but four or more were rated as unclear risk; and all other cases were rated as high risk of bias.⁴ This approach is similar to using scales that add up scores for multiple items to produce a total, which is discouraged in the Cochrane Handbook.¹⁶ The Handbook instead recommends an overall qualitative assessment considering the relative importance of different domains.¹⁶ According to our knowledge, none of their methods are supported by empirical evidence. The authors rated 96 (18%) of the 522 trials as low risk of bias, 380 (73%) trials as 'moderate', and 46 (9%) trials as high risk of bias. We were not able to replicate these findings and those efforts were made difficult because it was not clear how the blinding domains were rated in terms of risk of bias. Given that that the review's five outcomes were all likely affected by all of the risk of bias domains, the qualitative method suggested by the Cochrane Handbook involves classifying trials with any 'high risk of bias' domains as overall high risk of bias.¹⁶ Applying these criteria (Cochrane Handbook, Table 8.7.a)¹⁶ on Cipriani et al.'s ratings, there was one trial at low risk of bias, 383 trials (73%) at unclear risk, and 138 trials (26%) at high risk of bias. When we used our classifications for the blinding domains (i.e. all placebo-controlled trials were rated as unclear risk of bias, and for the selective outcome reporting domain) there were no (0%) trials at low risk of bias, 261 (50%) trials of unclear risk, and 261 (50%) trials of high risk of bias (S1 Appendix). If the three blinding domains were rated as high risk of bias in the placebo-controlled trials, rather than unclear risk of bias, there were no (0%) trials at low risk, 108 trials (21%) of unclear risk, and 414 trials (79%) of high risk of bias (S1 Appendix).

Publication bias

Publication bias of antidepressant trials is pervasive and distorts the evidence base.⁸ Many industry funded antidepressant trials remain unpublished or are inadequately reported.⁸ Cipriani et al.⁴ included 436 published and 86 unpublished studies, but as many as a thousand antidepressant studies may have been conducted.¹² We did a random effects meta-analysis of the placebo comparisons according to publication status and found that the average effect size was lower in unpublished studies (SMD 0.15 (95% CI: 0.11 to 0.19, 96 comparisons, 57 trials)) than in published studies (SMD 0.33 (95% CI: 0.30 to 0.35, 294 comparisons, 196 trials) (p<0.0001 for difference between the two estimates) (Table 1). Our findings are very similar to those reported by Turner et al.⁸ in 2008 of published versus unpublished antidepressant trials registered by the FDA who found an SMD of 0.37 (95% CI: 0.33 to 0.41) for published studies and 0.15 (95% CI: 0.08 to 0.22) for unpublished studies. This indicates that the reported effect sizes by Cipriani et al.⁴ are likely

1
2
3
4 inflated due to publication bias. They correctly downgraded their confidence in the evidence due to
5 the risk of publication bias, but it would also have been appropriate to estimate the impact of
6 publication bias on their effect estimate.
7

8 9 Trial duration and long-term effects

10 Cipriani et al.⁴ extracted outcome data as close to eight weeks follow-up as possible within an
11 interval of four to 12 weeks,⁴ but they did not provide a rationale for this decision.²¹ The common
12 clinical practice is to prescribe antidepressants for much longer periods. In the Netherlands, 43% of
13 SSRI users receive treatment for 15 months or more,²⁶ while 68% of those who use
14 antidepressants in the US take them for two years or more, and 25% take them for more than 10
15 years.¹ Although the short trial duration was acknowledged by the authors as a limitation, the lack
16 of clinical relevance of such short follow-up should have been highlighted and the confidence in the
17 evidence should have been downgraded one level in the GRADE domain of 'indirectness'. A more
18 appropriate method would have been to extract outcome data according to length of treatment and
19 follow-up to assess any change in the treatment effect over time. According to the trial
20 characteristics reported by Cipriani et al.⁴ 12 (4%) of the 304 placebo-controlled trials lasted more
21 than 12 weeks. However, that figure was misleading because only four of these 12 trials contained
22 an uninterrupted double-blind, placebo-controlled phase of more than 12 weeks (S2 Appendix).
23 The two placebo-controlled trials with the longest follow-up included 81 participants at 36 weeks
24 (S2 Appendix). A further consequence of a short follow-up period is an underestimation of serious
25 and non-serious adverse events.²⁷
26
27
28

29 30 'Placebo run-in' and inclusion of already treated patients

31 The 'placebo run-in' study design distorts the estimates of benefits and harms (Table 2A). Cipriani
32 et al. did not provide a clear definition of a 'placebo run-in',²¹ but they characterised 260 (50%) of
33 the 522 included trials as having a 'placebo run-in', 182 (35%) trials as unclear, and 80 (15%)
34 trials as having no 'placebo run-in'.⁴ We performed random effects meta-analyses of the placebo-
35 controlled trials according to the use of a 'placebo run-in' design and found that the effect sizes
36 differed between the groups with an SMD of 0.31 (95% CI: 0.28 to 0.34, 221 comparisons, 142
37 trials) in trials with a 'placebo run-in', an SMD of 0.29 (95% CI: 0.25 to 0.33, 120 comparisons, 79
38 trials) where the use of a 'placebo run-in' was unclear, and an SMD of 0.22 (95% CI: 0.16 to 0.29,
39 46 comparisons, 30 trials) in trials without a 'placebo run-in' ($p=0.05$ for the difference between the
40 three estimates). In a further subgroup analysis of unpublished trials without 'placebo run-in' the
41 effect size was very small (SMD 0.08, 95% CI: -0.27 to 0.11, 8 comparisons, 5 trials). The use of
42 the 'placebo run-in' design and its implications were not discussed by Cipriani et al.⁴
43
44
45
46

47 **Table 2.** 'Placebo run-in', minimal clinically significant difference, and 'response' as an outcome.

48 49 50	A. 'Placebo run-in' and the inclusion of already treated participants distort the benefit-harm balance
----------------	---

51 52 53 54 55	Cipriani et al. ⁴ did not provide a definition of 'placebo run-in', but it usually involves that the participants, before the randomisation, receive placebo, typically for about a week after which non-adherent participants and those who responded well to the placebo (often called "placebo-
----------------------------	---

responders”) are excluded from the trial. Participants already in treatment with antidepressants, including the study drug, are virtually always allowed to enter the trial, and commonly all participants are tapered off ongoing antidepressant medication during the ‘placebo run-in’. This study design may impact the effect estimates of placebo-controlled trials and the benefit/harm balance through several mechanisms that favour the drug over placebo:

- Participants treated with the study drug, or a similar drug, prior to inclusion and subsequently randomised to the drug will most likely tolerate it and experience fewer harms compared to a drug naïve population (**reduced harms in the drug group**).
- Participants treated with an antidepressant before the trial and subsequently randomised to placebo might experience withdrawal symptoms that can be misinterpreted as signs of worsening of the depression or as adverse events.²⁸ Withdrawal symptoms typically occur within a few days after discontinuation but there is great clinical variation²⁸ (**reduced benefits and increased harms in the placebo group**).
- Participants already treated with an antidepressant and subsequently randomised to the study drug might experience withdrawal symptoms during the placebo run-in that are alleviated by the study drug.²⁸ It could be misinterpreted as an improvement of the depression (**increased benefits in the drug group**).

B. ‘Response rates’ lack clinical meaning

The ‘response rate’ is usually defined as the number of participants in a randomised clinical trial who achieve a reduction of more than 50% of the total score on a standardised observer-rated scale for depression, such as the Hamilton depression rating scale or the Montgomery-Åsberg rating scale. ‘Non-response’ does not necessarily imply that the participant’s condition has not improved, but simply that the improvement is rated to be less than the 50% reduction. The difference might be as little as one point on the rating scale between a ‘responder’ and a ‘non-responder’. Thus, participants classified as ‘non-responders’ may actually have shown substantial improvement. The difference in ‘response rates’ between antidepressants and placebo does therefore not indicate the difference in the number of participants who have improved, but only the difference in the number of participants whose improvement exceeded the arbitrarily defined threshold. In addition, by focusing on the number of participants crossing the 50% reduction threshold the participants whose conditions deteriorate during the trial are ignored. Therefore, it seems more clinically meaningful to look at the average effect estimate of the drug compared to placebo.

C. Minimal clinically relevant difference

Cipriani et al. reported an overall effect estimate measured as a standardised mean difference of 0.3 between antidepressants and placebo.⁴ The British National Institute of Health and Clinical Excellence suggested in 2004 a difference of three points on the Hamilton depression rating scale, or a standardised mean difference of 0.5, as a clinically significant change.⁵ However, this difference was arbitrary and not based on empirical data.²⁹ Leucht et al. used clinical trial data in

2013 to suggest that clinicians are unable to detect reductions on the Hamilton depression rating scale of three points or less.³⁰ Others have interpreted the same data and suggested that changes of seven points or more on the Hamilton scale, corresponding to a standardised mean difference of at least 0.875, are necessary for a clinician to detect a minimal clinical improvement.³¹ We found that the mean difference between antidepressants and placebo on the 17-item Hamilton depression rating scale (range 0 to 52 points), based on Cipriani et al.'s data,⁴ was 1.97 points. This mean difference falls below what is considered a clinically relevant effect.

Dropout as a proxy for harms

Overall dropout rates and dropouts due to adverse effects were assessed by Cipriani et al. as measures of “acceptability” and “tolerability”, respectively, whereas the antidepressants’ actual harms and serious and non-serious adverse events were not assessed. It can be meaningful to use total dropout rates as a measure of the overall benefit/harm balance, but due to the biases introduced by including participants who are already known to tolerate an antidepressant drug and the use of a ‘placebo run-in’, this outcome will likely be biased in favour of the active drug (Table 2A). Furthermore, by not including a careful analysis of the serious harms, which include aggression, suicide and death,³² and of specific adverse events, the review provided no basis for balancing the benefits and harms, which is essential for informed consent and shared clinical decision-making and for evaluating the drugs’ clinical value. Adverse effects of antidepressants are common and a recent meta-analysis of 131 trials of SSRIs for depression found an increased risk of serious adverse events compared to placebo (OR 1.37; 95% CI: 1.08 to 1.75).⁶ This is likely an underestimate, as only 44 of the 131 included trials reported these data⁶ and as serious harms, including death, of antidepressants are often not reported in published papers.³³ Except for two drugs none of the included antidepressants had statistically significant lower total dropout rates than placebo.⁴ However, Cipriani et al. likely underestimated the antidepressants’ total dropout rates since they were missing in 58 (11%) of the trials and the dropout rates due to adverse events were missing in 93 (18%) of the trials. A meta-analysis of dropouts in 73 trials based on clinical study reports obtained from drug regulators, rather than published data, showed that 12% more participants dropped out on antidepressants than on placebo.³⁴

We had access to the clinical study reports for 19 of the 522 trials included in Cipriani et al.’s review. The outcomes of total dropout rates and dropout rates due to adverse events were fully reported in all 19 clinical study reports. In comparison with those data, total dropout rates or dropouts due to adverse events were either not reported or incorrectly reported by Cipriani et al. in 12 (63%) of the 19 trials: total dropout rates were not reported for two trials and incorrectly reported for seven trials; dropouts due to adverse events were not reported for five trials and incorrectly reported for three trials (S1 Table).

Lack of patient relevant outcomes

Patient relevant outcomes such as quality of life and sick leave are rarely measured and reported in psychiatric drug trials. Instead, the trials mostly rely on investigator-rated symptom scores, although self-rated symptom scales also exist. In a systematic review of SSRIs for depression in adults, only six of 131 trials reported quality of life data⁶ and even clinical study reports are

1
2
3
4 unreliable because of selective reporting of this outcome.³⁴ The inability to cope with daily activities
5 and the drugs' side-effects may be more important to patients than their depressed mood³⁵ and
6 the exclusion of patient-relevant outcomes in the protocol²¹ is a major limitation of the evidence
7 and of Cipriani et al.'s overall conclusion.⁴
8
9

10 Clinically irrelevant efficacy outcomes

11 The network meta-analysis' primary efficacy outcome was 'response rate' (Table 2B). It is a
12 problematic outcome because it lacks clinical relevance and it may create an illusion of clinical
13 effectiveness.³⁶ Dichotomisation of outcomes measured on rating scales leads to loss of statistical
14 power, and it increases the risk of false positive results³⁷ and spuriously inflated effect sizes.³⁶
15 Therefore, methodologists discourage the use of such dichotomised outcomes and they should
16 generally be avoided when rating scale data are available.³⁷ These issues also apply to the
17 review's secondary outcome of 'remission rates'. The choice made by Cipriani et al.⁴ to report only
18 the relative odds ratios and not the trials' absolute 'response rates' has been criticised.³⁸ However,
19 even the absolute 'response rates' are of limited clinical relevance. Cipriani et al.⁴ did not address
20 the problems related to 'response' and 'remission rates'.
21
22
23

24 Statistical versus clinical significance

25 Cipriani et al.⁴ also reported the standardised mean difference (SMD) on symptom rating scales,
26 which is more meaningful than the dichotomised outcomes.^{36 37} They reported an overall SMD for
27 antidepressants versus placebo of 0.30 (95% credible interval: 0.26 to 0.34), but the number of
28 trials and comparisons were unclear.⁵ We found a similar overall SMD for antidepressants versus
29 placebo for the direct pair-wise comparisons of 0.29 (95% CI: 0.27 to 0.31, 390 comparisons, 253
30 studies) (Table 1). These effect estimates are statistically significant, but below what is considered
31 a clinically relevant effect (Table 2C). We also calculated an overall mean difference for the trials
32 that reported endpoint or change scores on the 17-item Hamilton depression rating scale, which
33 was the most commonly used scale in the included trials (S2 Table). The mean difference between
34 antidepressants and placebo was 1.97 points (95% CI: 1.74 to 2.21, 166 comparisons, 109 trials)
35 on the 17-item Hamilton depression rating scale (range 0 to 52) (Table 1). This mean difference on
36 the Hamilton scale also falls below what is considered a clinically relevant effect (Table 2C).
37 Cipriani et al. did not discuss the clinical significance of their reported effect size.⁴
38
39
40

41 Selected, non-representative study populations

42 Antidepressant trials typically have extensive exclusion criteria that limit their external validity.
43 These include psychiatric comorbidities, alcohol abuse, long duration of illness, and 'non-response'
44 to previous antidepressant treatment.³⁹ The majority of patients in a clinical setting would not be
45 eligible to enter randomised trials due to such exclusion criteria,⁴⁰ and the evidence coming from
46 these trials is therefore of limited relevance. Furthermore, the exclusion of previous 'non-
47 responders' and inclusion of those who are expected to respond more favourably to treatment may
48 bias the trials (Table 2A). These issues were not considered by Cipriani et al.⁴ but should arguably
49 have resulted in downgrading of the confidence in the evidence in the GRADE domain of
50 indirectness.¹⁷
51
52
53

54 The certainty of the evidence

55
56
57
58
59
60

1
2
3
4 Cipriani et al.⁴ assessed the certainty of evidence for the two main outcomes using the GRADE
5 approach adapted for network meta-analyses. They provided the GRADE results for the head-to-
6 head comparisons, but we were unable to find the results for the placebo comparisons.⁴
7 Following the issues related to the quality of the evidence, the certainty of evidence for the placebo
8 comparisons should arguably be downgraded two levels due to a 'high risk' of bias, two levels in
9 the domain of 'indirectness' due to short trial lengths, strict inclusion criteria and the use of 'placebo
10 run-in', in addition to downgrading one level due to publication bias as acknowledged by Cipriani et
11 al.⁴ Downgrading due to the indirectness of the network meta-analysis' methodology should also
12 be considered.⁴¹ Taken together, the certainty of evidence should be 'very low'.¹⁷
13
14

15 Discussion

16 We have identified several important biases that were not taken into account in the systematic
17 review by Cipriani et al.⁴ We showed that the reported effect of antidepressants over placebo
18 measured on depression rating scales was small and likely inflated by several methodological
19 flaws in the trials. For the first time, we showed that the 'placebo run-in' study design appears to
20 work towards producing inflated effect sizes, in addition to publication bias and other flaws.
21 Further, we showed that the outcome data reported by Cipriani et al. differed from the clinical study
22 reports and that their risk of bias assessment did not follow the methods outlined in the Cochrane
23 Handbook. Finally, we found that the certainty of evidence for antidepressants versus placebo for
24 all outcomes assessed should be very low. Taken together, the evidence does not support
25 definitive conclusions regarding the efficacy of antidepressants for depression in adults, including
26 whether they are more efficacious than placebo for depression.
27
28
29

30 Previous meta-analyses (Figure 1) have found similar improvement in symptom scores as Cipriani
31 et al.⁴ Several of these reviews have considered carefully the methodological flaws, assessed the
32 harms, and have drawn different conclusions.⁵⁻⁷ We found that Cipriani et al. did not assess the
33 risk of bias in accordance with the Cochrane Handbook as stated⁴ and their results were non-
34 transparently presented. While the authors should be commended for sharing their data, most of
35 the review's results cannot be reproduced because basic information, such as the number of
36 included studies, arms and participants for each meta-analysis, was not reported. The network
37 meta-analysis methodology may hold some promise, but only in areas where clearly effective
38 interventions exist and need to be ranked., and the many statistical options should never
39 overshadow an initial critical assessment of the evidence and a clear presentation of the results. It
40 seems misleading to rank the antidepressants when we have very low confidence in the evidence.
41 Interestingly, our pairwise meta-analysis of improvement on symptom scores yielded very similar
42 results to those reported by Cipriani et al. The added benefit of the network meta-analysis
43 methodology therefore seems unclear.⁴
44
45
46

47 We found that the evidence base consists of mainly short-term trials (12 weeks or less) with no
48 evidence for treatment beyond 36 weeks although most patients are treated for years.^{1,26} Further,
49 the apparent effect of antidepressants reported in the review by Cipriani et al.⁴ measured on
50 investigator-rated symptom scales was small and likely not clinically relevant. Observational
51 studies also indicate that the effectiveness of antidepressants in practice is very low. In the large,
52 publicly funded, Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, only
53 3% of the 4041 enrolled patients were considered "in remission" after one year.⁴² The recent
54 finding based on clinical study reports of randomised trials that more participants drop out on
55
56
57
58
59
60

1
2
3
4 antidepressants than on placebo,³⁴ further suggests that the benefits of antidepressants may not
5 outweigh the harms.
6

7 Our findings showed that Cipriani et al.'s data⁴ were inaccurate, and their estimates may therefore
8 be incorrect because they relied on published data. It may be perceived as a limitation that we
9 relied on the data by Cipriani et al. and did not perform our own separate systematic literature
10 search and data extraction. Considering the multiple methodological flaws we have identified, it
11 would be necessary to analyse data based on clinical study reports and individual patient data to
12 make a reliable assessment of the benefits and harms of antidepressants because they are the
13 most reliable source of trial data.⁴³ There are also some limitations to our sponsorship subgroup
14 analysis: While industry sponsored studies have been found to report favourable efficacy results
15 more often than non-industry sponsored studies,²⁵ our analysis showed that industry sponsored
16 trials reported a lower effect estimate of antidepressants compared to placebo than non-industry
17 sponsored trials on investigator-rated depression symptom scales. However, there were important
18 differences between the two subgroups that likely contributed to the observed difference (S1
19 Figure). Non-industry sponsored trials were smaller and older than industry sponsored trials and
20 almost all of the non-industry sponsored trials included by Cipriani et al. were published.
21
22
23

24 Our results highlight that the many hundreds of placebo-controlled trials of antidepressants have
25 not addressed the most important, patient-relevant questions regarding antidepressants' benefits
26 and harms. Although this has been known for years,¹² it has not led to changes in research
27 practice. Erroneous conclusions that antidepressants are efficacious for depression have the effect
28 that they may prevent people suffering from depression from seeking other solutions to alleviate
29 their condition, such as psychotherapy and dealing with psychosocial stressors, and they may stall
30 funding and research of such treatment modalities. Importantly, such conclusions may also lead to
31 a loss of interest in providing a better evidence base to determine the true clinical value of
32 antidepressants.
33
34

35 Our review has two implications. First, the review by Cipriani et al.⁴ and its conclusion should be
36 carefully revisited. In the light of our findings the review should not inform clinical practice. Second,
37 our reanalysis has highlighted the need for a radical change in the way antidepressant trials are
38 being conducted, reported, and interpreted. We hope that doctors, patients, peers, and politicians
39 will consider the limitations of the current evidence of antidepressants for depression that we have
40 presented and collectively act accordingly. This involves informing the patients about the limitations
41 of the current evidence, thus providing a basis for a true informed consent, and working toward a
42 better evidence base for the use of antidepressants in the treatment of depression. To get reliable
43 answers about the antidepressants' benefits and harms in adults with depression we need large-
44 scale, industry-independent, and better blinded, long-term trials of drug naïve participants, with
45 patient-relevant outcomes rather than ranking scales.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure legends

Figure 1. Previous meta-analyses reporting effect sizes for antidepressants versus placebo in adults.

Data are reported as standardised mean differences with 95% confidence intervals.

NICE 2004⁵: SSRIs. Kirsch 2008⁷: 'New generation' antidepressants. Turner 2008⁸: All antidepressants. Arroll 2009⁴⁴: Antidepressants for depression in primary care. Data represent a pooled estimate of tricyclic antidepressants and SSRIs versus placebo, fixed effects model. Fournier 2010⁴⁵: All antidepressants. Data represent pooled estimate from three groups of severity (mild to moderate, severe, very severe), fixed effects model. Gibbons 2012⁴⁶: Fluoxetine and venlafaxine. Jakobsen 2017⁶: SSRIs. The effect size of mean change scores. Cipriani 2018⁴: All antidepressants.

For peer review only

Contributors

KM, ASP-M and KB contributed to the conception and design of the study. KM performed the meta-analyses. KM, ASP-M and KB analysed and interpreted the data, drafted and critically revised the manuscript and approved the final version to be published. All authors had full access to all the data and take responsibility for the integrity of the data and the accuracy of the data analysis.

Data sharing

All data files and the code for the statistical analyses are available from the Open Science Framework database (Access for peer review: https://osf.io/3prz9/?view_only=45a0afe9a09d4e9e9eb8a2c2d8d4bdc6. A DOI will be assigned after acceptance of the manuscript for publication).

Funding

The study was funded by the Nordic Cochrane Centre. The funder had no role in the design, data collection, analysis, interpretation of data, writing of the report, or in the decision to submit the article for publication. The research was designed, conducted, analysed, and interpreted by the authors entirely independently of the funding source.

Ethical approval

Not required.

Competing interests

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

Acknowledgments

We are grateful to Luis Carlos Saiz, Pharm D, PhD for valuable comments.

References

1. U.S. Department of Health and Human Services: Centers for Disease Control and Prevention. Antidepressant Use Among Persons Aged 12 and Over: United States, 2011–2014 2017 [Available from: <https://www.cdc.gov/nchs/data/databriefs/db283.pdf> accessed 1 May 2018].
2. Danish Health Authority. MedicinForbrug - Indblik 2017: Laveste antal brugere af antidepressiv medicin de seneste 10 år [Medicine usage - 2017: Lowest number of antidepressant users in the last 10 years] 2017 [Available from: [https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/laegemidler/analyser-om-antidepressiva/laveste-antal-brugere-af-antidepressiv-medicin-de-seneste-10-aar-\(pdf\)-2017.pdf?la=da](https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/laegemidler/analyser-om-antidepressiva/laveste-antal-brugere-af-antidepressiv-medicin-de-seneste-10-aar-(pdf)-2017.pdf?la=da) accessed 1 May 2018].
3. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord* 2018 doi: 10.1111/bdi.12609 [published Online First: 2018/03/15]
4. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
5. National Institute for Health and Care Excellence. Depression: Management of Depression in Primary and Secondary Care. Clinical Practice Guideline Number 23. London: National Institute for Health and Clinical Excellence; 2004 [Available from: <http://webarchive.nationalarchives.gov.uk/20080612001911/https://www.nice.org.uk/guidance/index.jsp?action=download&o=29617> accessed 1 May 2018].
6. Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC psychiatry* 2017;17(1):58. doi: 10.1186/s12888-016-1173-2 [published Online First: 2017/02/10]
7. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5(2):e45. doi: 10.1371/journal.pmed.0050045 [published Online First: 2008/02/29]
8. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358(3):252-60. doi: 10.1056/NEJMsa065779 [published Online First: 2008/01/18]
9. The Guardian. The drugs do work: Antidepressants are effective, study shows: The Guardian; [updated 21 February 2018. Available from: <https://www.theguardian.com/science/2018/feb/21/the-drugs-do-work-antidepressants-are-effective-study-shows> accessed 1 May 2018].
10. TIME. These antidepressants are most effective, study says Time Magazine: Time Magazine; [updated 21 February 2018. Available from: <http://time.com/5169013/antidepressants-more-effective-placebo-treating-depression/> accessed 1 May 2018].
11. Hengartner MP. Methodological flaws, conflicts of interest, and scientific fallacies: Implications for the evaluation of antidepressants' efficacy and harm. *Front Psychiatry* 2017;8:275. doi: 10.3389/fpsy.2017.00275 [published Online First: 2017/12/23]
12. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14. doi: 10.1186/1747-5341-3-14 [published Online First: 2008/05/29]
13. Boesen K, Paludan-Müller AS, Munkholm K. Network meta-analysis, antidepressants and misleading bias assessments. *The Lancet (accepted for publication)* 2018;392
14. Gotzsche PC, Jorgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686. doi: 10.1136/bmj.d2686 [published Online First: 2011/05/12]
15. Int'Hout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard

- 1
2
3
4 DerSimonian-Laird method. *BMC medical research methodology* 2014;14:25. doi:
5 10.1186/1471-2288-14-25 [published Online First: 2014/02/20]
6 16. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions
7 Version 5.1.0 [updated March 2011]. Available from <http://handbook-5-1cochraneorg/>
8 (accessed 1 May 2018): The Cochrane Collaboration, 2011.
9 17. Schünemann H, Brożek J, Guyatt G, et al. GRADE handbook for grading quality of evidence
10 and strength of recommendations. Updated October 2013. Available from
11 guidelinedevelopment.org/handbook2013.
12 18. Page MJ, Higgins JP, Clayton G, et al. Empirical evidence of study design biases in
13 randomized trials: Systematic review of meta-epidemiological studies. *PLoS One*
14 2016;11(7):e0159267. doi: 10.1371/journal.pone.0159267 [published Online First:
15 2016/07/12]
16 19. Perlis RH, Ostacher M, Fava M, et al. Assuring that double-blind is blind. *Am J Psychiatry*
17 2010;167(3):250-2. doi: 10.1176/appi.ajp.2009.09060820 [published Online First:
18 2010/03/03]
19 20. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression.
20 *Cochrane Database Syst Rev* 2004(1):CD003012. doi: 10.1002/14651858.CD003012.pub2
21 [published Online First: 2004/02/20]
22 21. Furukawa TA, Salanti G, Atkinson LZ, et al. Comparative efficacy and acceptability of first-
23 generation and second-generation antidepressants in the acute treatment of major
24 depression: protocol for a network meta-analysis. *BMJ Open* 2016;6(7):e010919. doi:
25 10.1136/bmjopen-2015-010919 [published Online First: 2016/07/13]
26 22. Woolley SB, Cardoni AA, Goethe JW. Last-observation-carried-forward imputation method in
27 clinical efficacy trials: review of 352 antidepressant studies. *Pharmacotherapy*
28 2009;29(12):1408-16. doi: 10.1592/phco.29.12.1408 [published Online First: 2009/12/02]
29 23. Dziura JD, Post LA, Zhao Q, et al. Strategies for dealing with missing data in clinical trials: from
30 design to analysis. *Yale J Biol Med* 2013;86(3):343-58. [published Online First: 2013/09/24]
31 24. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of
32 outcomes in randomized trials: comparison of protocols to published articles. *JAMA*
33 2004;291(20):2457-65. doi: 10.1001/jama.291.20.2457 [published Online First: 2004/05/27]
34 25. Lundh A, Lexchin J, Mintzes B, et al. Industry sponsorship and research outcome. *Cochrane*
35 *Database Syst Rev* 2017;2:MR000033. doi: 10.1002/14651858.MR000033.pub3 [published
36 Online First: 2017/02/17]
37 26. Huijbregts KM, Hoogendoorn A, Slottje P, et al. Long-term and short-term antidepressant use
38 in general practice: Data from a large cohort in the netherlands. *Psychother Psychosom*
39 2017;86(6):362-69. doi: 10.1159/000480456 [published Online First: 2017/11/14]
40 27. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and
41 opportunities. *Trials* 2012;13:138. doi: 10.1186/1745-6215-13-138 [published Online First:
42 2012/08/22]
43 28. Fava GA, Gatti A, Belaise C, et al. Withdrawal symptoms after selective serotonin reuptake
44 inhibitor discontinuation: A systematic review. *Psychother Psychosom* 2015;84(2):72-81.
45 doi: 10.1159/000370338 [published Online First: 2015/02/28]
46 29. Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336(7643):516-7. doi:
47 10.1136/bmj.39510.531597.80 [published Online First: 2008/03/06]
48 30. Leucht S, Fennema H, Engel R, et al. What does the HAMD mean? *J Affect Disord*
49 2013;148(2-3):243-8. doi: 10.1016/j.jad.2012.12.001 [published Online First: 2013/01/30]
50 31. Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of
51 antidepressant-placebo differences. *Contemp Clin Trials* 2015;43:60-2. doi:
52 10.1016/j.cct.2015.05.005 [published Online First: 2015/05/17]
53 32. Sharma T, Guski LS, Freund N, et al. Suicidality and aggression during antidepressant
54 treatment: systematic review and meta-analyses based on clinical study reports. *BMJ*
55 2016;352:i65. doi: 10.1136/bmj.i65 [published Online First: 2016/01/29]
56
57
58
59
60

- 1
2
3
4 33. Hughes S, Cohen D, Jaggi R. Differences in reporting serious adverse events in industry
5 sponsored clinical trial registries and journal articles on antidepressant and antipsychotic
6 drugs: a cross-sectional study. *BMJ Open* 2014;4(7):e005535. doi: 10.1136/bmjopen-2014-
7 005535 [published Online First: 2014/07/11]
- 8 34. Sharma T. Effects of selective serotonin reuptake inhibitors (SSRIs) and serotonin -
9 norepinephrine reuptake inhibitors (SNRIs) on suicidality, violence, and quality of life (PhD
10 Thesis). Available from:
11 [http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_ph](http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_phd.pdf)
12 [d.pdf](http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_phd.pdf). University of Copenhagen, 2018.
- 13 35. Eiring O, Landmark BF, Aas E, et al. What matters to patients? A systematic review of
14 preferences for medication-associated outcomes in mental disorders. *BMJ Open*
15 2015;5(4):e007848. doi: 10.1136/bmjopen-2015-007848 [published Online First:
16 2015/04/10]
- 17 36. Kirsch I, Moncrieff J. Clinical trials and the response rate illusion. *Contemp Clin Trials*
18 2007;28(4):348-51. doi: 10.1016/j.cct.2006.10.012 [published Online First: 2006/12/22]
- 19 37. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*
20 2006;332(7549):1080. doi: 10.1136/bmj.332.7549.1080 [published Online First: 2006/05/06]
- 21 38. McCormack J, Korownyk C. Effectiveness of antidepressants. *BMJ* 2018;360:k1073. doi:
22 10.1136/bmj.k1073 [published Online First: 2018/03/11]
- 23 39. Posternak MA, Zimmerman M, Keitner GI, et al. A reevaluation of the exclusion criteria used in
24 antidepressant efficacy trials. *Am J Psychiatry* 2002;159(2):191-200. doi:
25 10.1176/appi.ajp.159.2.191 [published Online First: 2002/02/02]
- 26 40. Zetin M, Hoepner CT. Relevance of exclusion criteria in antidepressant clinical trials: a
27 replication study. *J Clin Psychopharmacol* 2007;27(3):295-301. doi:
28 10.1097/JCP.0b013e318058263f [published Online First: 2007/05/16]
- 29 41. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--
30 indirectness. *J Clin Epidemiol* 2011;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014
31 [published Online First: 2011/08/02]
- 32 42. Pigott HE. The star*d trial: It is time to reexamine the clinical beliefs that guide the treatment of
33 major depression. *Can J Psychiatry* 2015;60(1):9-13. doi: 10.1177/070674371506000104
34 [published Online First: 2015/04/18]
- 35 43. Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports:
36 recommendations from the Tamiflu experience. *PLoS Med* 2012;9(4):e1001201. doi:
37 10.1371/journal.pmed.1001201 [published Online First: 2012/04/17]
- 38 44. Arroll B, Elley CR, Fishman T, et al. Antidepressants versus placebo for depression in primary
39 care. *Cochrane Database Syst Rev* 2009(3):CD007954. doi: 10.1002/14651858.CD007954
40 [published Online First: 2009/07/10]
- 41 45. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression
42 severity: a patient-level meta-analysis. *JAMA* 2010;303(1):47-53. doi:
43 10.1001/jama.2009.1943 [published Online First: 2010/01/07]
- 44 46. Gibbons RD, Hur K, Brown CH, et al. Benefits from antidepressants: synthesis of 6-week
45 patient-level outcomes from double-blind placebo-controlled randomized trials of fluoxetine
46 and venlafaxine. *Arch Gen Psychiatry* 2012;69(6):572-9. doi:
47 10.1001/archgenpsychiatry.2011.2044 [published Online First: 2012/03/07]
- 48
49
50
51
52
53
54
55
56
57
58
59

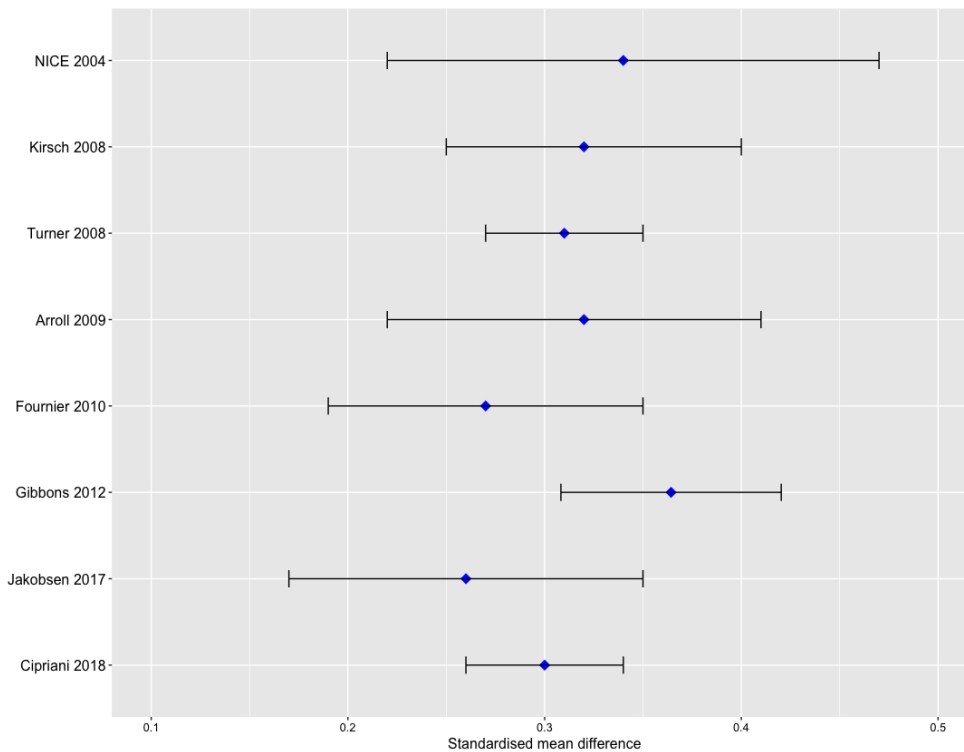


Figure 1. Previous meta-analyses reporting effect sizes for antidepressants versus placebo in adults.

Data are reported as standardised mean differences with 95% confidence intervals. NICE 2004⁵: SSRIs.

Kirsch 2008⁷: 'New generation' antidepressants. Turner 2008⁸: All antidepressants. Arroll 2009⁴⁴: Antidepressants for depression in primary care. Data represent a pooled estimate of tricyclic antidepressants and SSRIs versus placebo, fixed effects model. Fournier 2010⁴⁵: All antidepressants. Data represent pooled estimate from three groups of severity (mild to moderate, severe, very severe), fixed effects model.

Gibbons 2012⁴⁶: Fluoxetine and venlafaxine. Jakobsen 2017⁶: SSRIs. The effect size of mean change scores. Cipriani 2018⁴: All antidepressants.

389x293mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

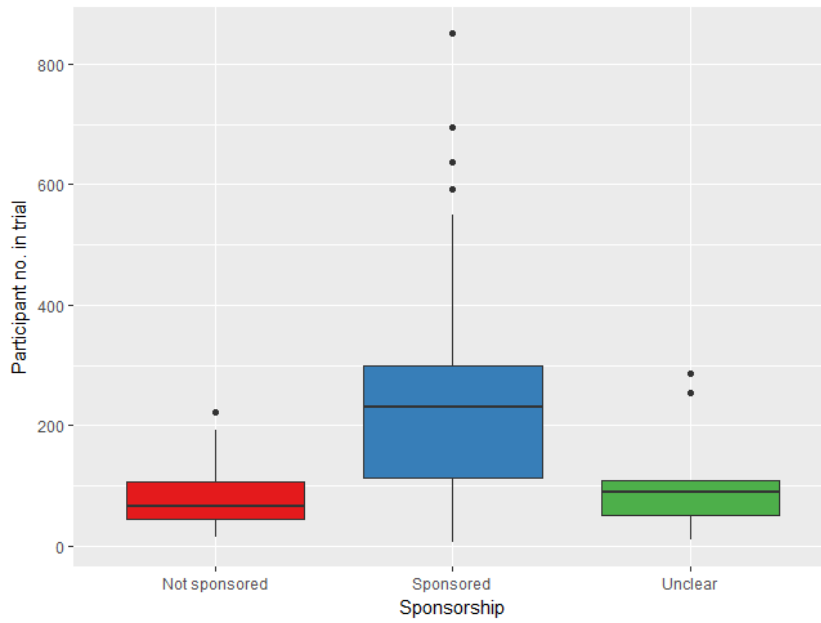
Considering the flaws in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

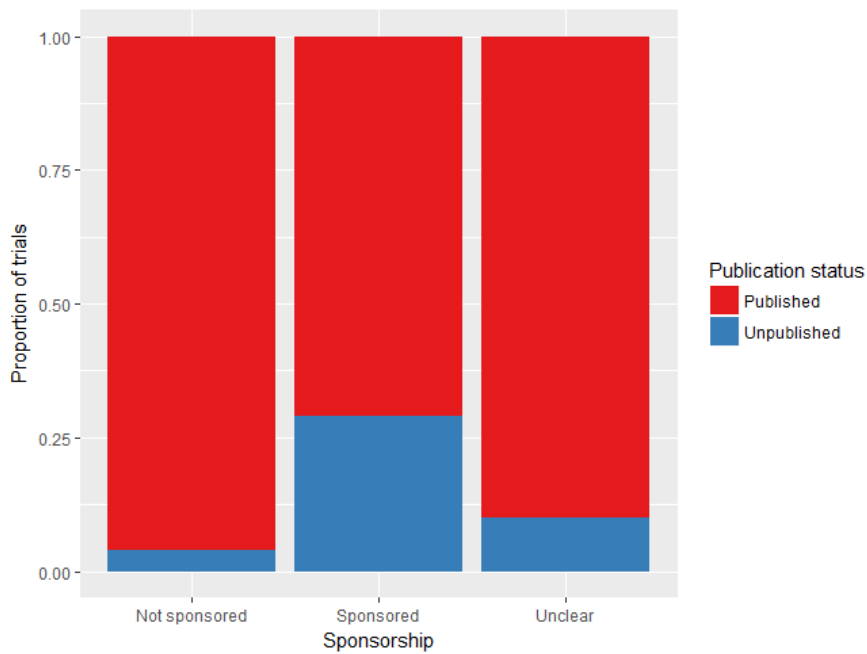
^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Figure

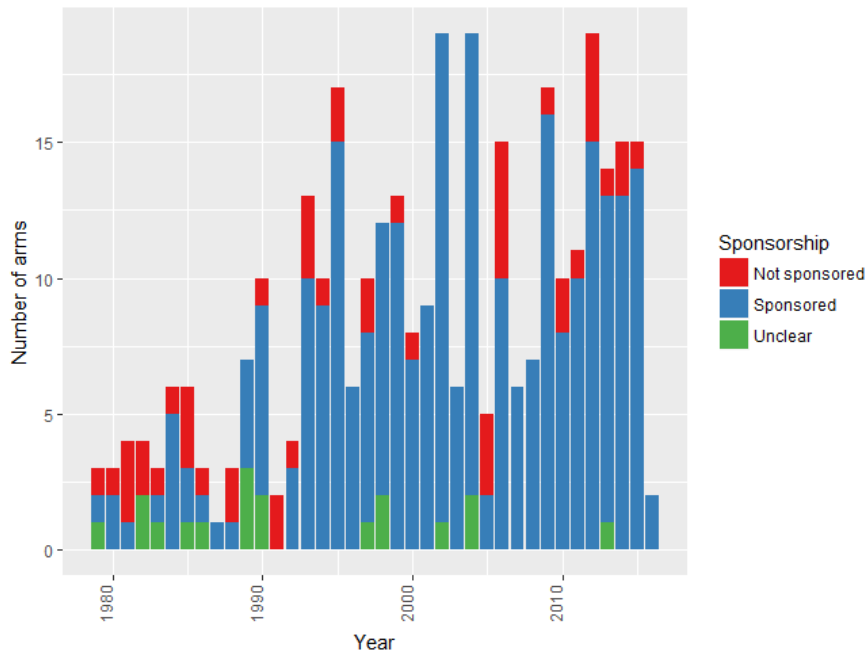
For peer review only



S1 Figure. A Size of placebo-controlled trials in the review by Cipriani et al.¹ according to sponsorship.



S1 Figure. Sponsorship status of placebo-controlled trials in the review by Cipriani et al.¹



S1 Figure. C Sponsorship of placebo-controlled trials in the review by Cipriani et al.¹ according to publication year.

Peer review only

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considering the flaws in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Table

For peer review only

S1 Table. Comparison between the outcomes of total dropout rates and dropouts due to adverse events as reported by Cipriani et al.¹ and the data reported in the clinical study reports.

Trial ID (From Cipriani et al. dataset)	Total dropouts reported by Cipriani et al. compared with clinical study reports*	Dropouts due to adverse events reported by Cipriani et al. compared with clinical study reports**
Goldstein2002 (HMAQ - Study Group A)	Match	Match
Study F1J-MC-HMAQ - Study Group B (starting page in the pdf document: 147)	Match	Not reported
Goldstein2004a (HMAQ - Study Group A, ID#4091)	Discrepancy	Match
Goldstein2004b (HMAQ - Study Group B, ID#4091)	Match	Discrepancy
Detke2004 (HMAQ Study Group A)	Match	Match
Perahia2006 (HMAQ - Study Group B)	Match	Match
Detke2002a (HMBH - Study Group A)	Match	Match
Detke2002b (HMBH - Study Group B)	Match	Discrepancy
0600B-367	Not reported	Not reported
0600B1-384	Discrepancy	Discrepancy
VEN 600A-303 (FDA)	Discrepancy	Match
Cunningham1994 (VEN 600A-302 FDA)	Discrepancy	Match
VEN 600A-313 (FDA)	Discrepancy	Match
0600A1-372	Not reported	Not reported
Thase1997 (VEN XR 209 FDA)	Match	Match
Rudolph1998 (VEN 600A-203 (FDA)	Match	Match
SER 101 (FDA)	Match	Not reported
Fabre1995 (SER 103 FDA)	Match	Match
SER 310 (FDA)	Discrepancy	Not reported

*Our comparison considered all treatment arms in the trials. If data for all arms matched, we assigned an assessment of 'match'; if data for any arm differed but were reported, we assigned an assessment of 'discrepancy'; if data for any arm was not reported, we assigned an assessment of 'not reported'.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]

For peer review only

1
2
3
4
5
6
7
8
9

Considering the flaws in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

10 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

11
12 ^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
13
14
15

16
17 S2 Table
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

S2 Table. Rating scales used in the included placebo-controlled trials (n = 304).¹

Scale	Number of trials	Number of treatment arms
HAMD17	137	198
HAMD21	92	139
HAMD24	13	22
HAMD29	1	2
HAMD31	1	2
HAMD unspecified	22	25
IDS-IVR-30	1	1
MADRS	29	50
NA	8	12

NA: not available; HAMD17: 17-item Hamilton depression rating scale; IDS-IVR-30: 30-item Inventory of Depressive Symptomatology-Self Report; MADRS: Montgomery-Åsberg Depression Rating Scale.

For peer review only

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considering the flaws in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Appendix

For peer review only

Risk of bias

Blinding of the participants

Cipriani et al.¹ classified three²⁻⁴ out of the 522 trials as being at low risk of bias in the “blinding of participants” domain:

Brunoni et al.² tested the blinding, by asking the participants who completed the trial to guess their allocation. They reported that 39 (75%) of 52 participants on placebo, and 29 (58%) of 50 participants on sertraline were able to correctly guess their allocation. Brunoni et al. suggested that the results were “driven by clinical improvement... rather than blinding failure”². We disagree and would categorise this trial at high risk of blinding bias of the participants.

Edwards et al.³ did not describe how and when they tested the blinding, and two of the included participants were not asked. They reported that 12 (60%) of 20 participants on placebo and 12 (63%) of 19 participants on paroxetine were able to correctly guess their treatment allocation. Edwards et al. concluded that their results “confirmed the blindness of the study”, but we would categorise the trial to be at unclear risk of bias.

Schatzberg et al.⁴ did not test the blinding, and it is unclear why this trial was rated at low risk of bias, rather than the “stated but not tested” categorisation.

Other bias domain

While Cipriani et al.¹ categorised each arm individually according to sponsorship, we considered sponsorship on the study level: trials with any sponsored arm (as categorised by Cipriani et al.¹) were categorised as “sponsored”; of the remaining trials, those with any arms categorised as “unclear” we labelled “unclear” and the remaining trials were categorised “not-sponsored”.

Summary risk of bias assessments

Criteria for assessments

To categorise the 522 trials included by Cipriani et al.¹, we followed the Cochrane Handbook’s criteria for an overall risk of bias assessment.⁵ Each domain in the risk of bias tool likely affects all five included outcomes assessed by Cipriani et al., and we therefore considered all bias domains as “key domains”, according to table 1.

Table 1. Criteria for overall risk of bias assessment.

Low risk of bias	All key domains classified as low risk of bias
Unclear risk of bias	One or more key domains classified as unclear risk of bias, and no domains classified as high risk of bias.
High risk of bias	One or more key domains classified as high risk of bias

We collapsed the three blinding domains in our Excel dataset and used the following criteria for our categorisation: All placebo-controlled trials were classified as unclear risk of bias in the main analysis and as high risk of bias in the sensitivity analysis. Trials that only contained head-to-head antidepressant arms, and no placebo arm, were rated as low risk of bias if the three blinding domains were rated as ‘low’ or ‘stated but not tested’ by Cipriani et al.¹. We rated the collapsed blinding domain as unclear risk of bias, if one or more of the blinding domains were rated as unclear by Cipriani et al. Trials with missing data for any of the five included outcomes were categorised as high risk of bias. For the remaining bias domains, we adopted the categorisations by Cipriani et al.¹. Our results are compared with Cipriani et al.’s in table 2.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 2. Comparison of the overall risk of bias assessments.

Cipriani et al. overall assessment		Our overall assessments			
		Cochrane categorisation	Sensitivity analysis using the Cipriani et al. categorisation	Our assessment	Sensitivity analysis of the blinding domains
Low risk	46 trials (9%)	Low	1 trial (0.2%)	0 trials (0%)	0 trials (0%)
"Moderate" risk	380 trials (73%)	Unclear	383 trials (73%)	261 trials (50%)	108 (21%)
High risk	96 trials (18%)	High	138 trials (26%)	261 trials (50%)	414 (79%)

For peer review only

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
2. Brunoni AR, Valiengo L, Baccaro A, et al. The sertraline vs. electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial. *JAMA psychiatry* 2013;70(4):383-91. doi: 10.1001/2013.jamapsychiatry.32 [published Online First: 2013/02/08]
3. Edwards JG, Goldie A. Placebo-Controlled trial of paroxetine in depressive illness. *Human Psychopharmacology: Clinical and Experimental* 1993;8(3):203-09. doi: 10.1002/hup.470080306
4. Schatzberg AF, Kremer C, Rodrigues HE, et al. Double-blind, randomized comparison of mirtazapine and paroxetine in elderly depressed patients. *Am J Geriatr Psychiatry* 2002;10(5):541-50. [published Online First: 2002/09/06]
5. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Available from <http://handbook-5-1.cochrane.org/> (accessed 1 May 2018): The Cochrane Collaboration, 2011.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considering the flaws in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S2 Appendix

For peer review only

Trial duration and long-term effect estimates

Overall trial length

According to the trial characteristics by Cipriani et al.¹ 492 (94%) of the 522 included trials lasted between four and 12 weeks (figure 1), 28 trials (5%) lasted more than 12 weeks, and the trial duration was unclear for two trials (table 3).

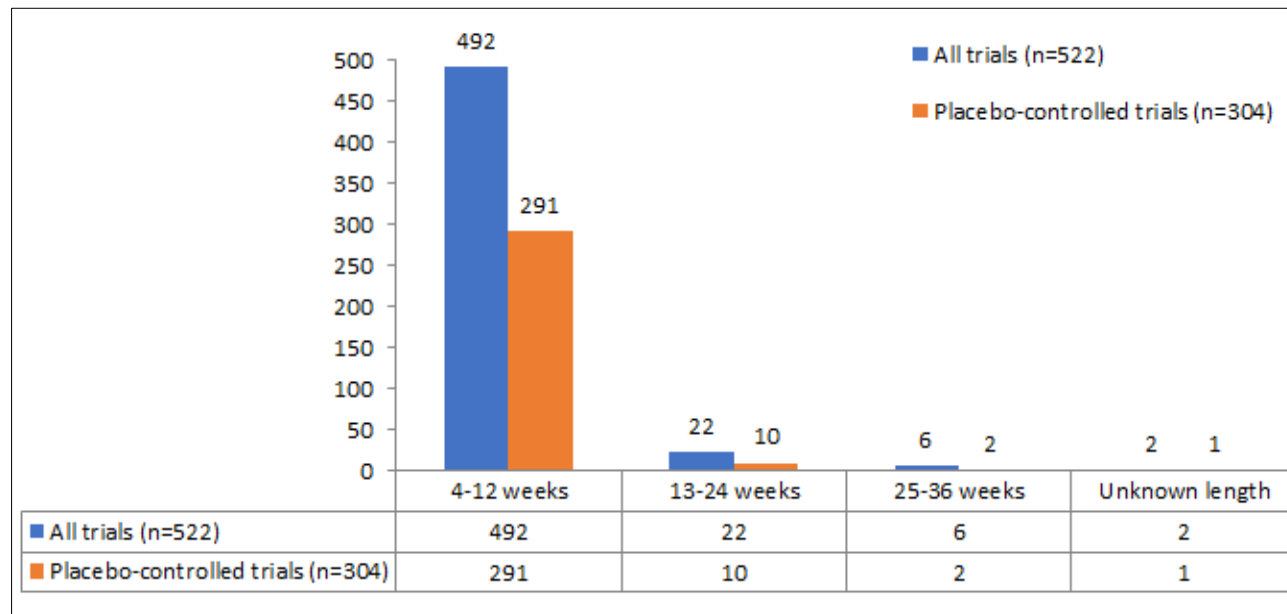


Figure 1. All 522 trials arranged into three follow-up periods according to Cipriani et al.’s trial characteristics: Four to 12 weeks (the period used by Cipriani et al.), 13 to 24 weeks, and 25 to 36 weeks ¹.

Longer-term placebo-controlled trials

Of the 28 trials lasting more than 12 weeks, 12 trials had a placebo-controlled arm (table 3) and 16 were head-to-head trials. However, upon closer examination, eight of the 12 placebo-controlled trials consisted of several ‘phases’ and only four trials contained a continuous randomised placebo-controlled phase of more than 12 weeks (table 4). It is unclear why Cipriani et al.¹ did not list the correct trial length characteristics, since the various extension phases were clearly described in the available documents, also for the unpublished trials.

Table 3. All 522 trials arranged according to the trial length characteristics by Cipriani et al. ¹.

Trial length according to Cipriani et al. ¹ (N = 522)	Trials without a placebo arm (N)	Placebo-controlled trials (N)
4 weeks	9	39
5 weeks	11	4
6 weeks	92	100
7 weeks	9	0
8 weeks	59	118
9 weeks	3	2
10 weeks	4	10
12 weeks	14	18
13 weeks	1	1
16 weeks	2	2
24 weeks	0	1

25 weeks	9	6
26 weeks	3	0
36 weeks	1	0
Trial length unknown	0	2
Total	218	304

Table 4. Longer-term placebo-controlled trials (trial length of more than 12 weeks).

Trial	Drug arm 1	Drug arm 2	Trial length according to Cipriani et al. (weeks)	Actual trial length (weeks)	Study design and phases
Oakes 2012a ²	duloxetine	-	36	36	-
Oakes 2012b ²	duloxetine	-	36	36	-
CL3-20098-022 ³	agomelatine	fluoxetine	24	6 + 18	“Optional double-blind placebo-controlled extension” (page 24) for “responders to treatment” (page 26)
CL3-20098-023 ³	agomelatine	paroxetine	24	6 + 18	“Optional double-blind placebo-controlled extension” (page 24) for “W6 responders” (page 26)
CL3-20098-024 ³	agomelatine	fluoxetine	24	6 + 18	“Optional double-blind placebo-controlled extension” (page 24) for “responders to treatment at W6” (page 26)
CL3-20098-026 ³	agomelatine	-	24	6 + 18	“18-week extension period” of “responders at W6” (page 42)
CL3-20098-070 ⁴	agomelatine	-	24	8 + 16	16-week extension period for responders
Robinson 2014 ⁵	duloxetine	-	24	12 + 12	From week 12 to 20 placebo rescue or dose increase were available
Lopez-Rodriguez 2004 ⁶	fluoxetine	-	20	20	-
Barber 2011 ⁷	sertraline	-	16	8 + 8	Non-responders on sertraline were switched to venlafaxin ER at week 8
Dimidjian 2006 ⁸	paroxetine	-	16	8 + 8	The blind was broken at week 8 and the placebo arm was offered other treatments
Lecrubier 1997 ⁹	venlafaxine	-	13	13	-

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
2. Oakes TM, Myers AL, Marangell LB, et al. Assessment of depressive symptoms and functional outcomes in patients with major depressive disorder treated with duloxetine versus placebo: primary outcomes from two trials conducted under the same protocol. *Hum Psychopharmacol* 2012;27(1):47-56. doi: 10.1002/hup.1262 [published Online First: 2012/01/14]
3. CHMP assessment report for Thymanax. EMEA/H/C/000916. 2008 [Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/000916/WC500038315.pdf accessed 1 May 2018.
4. EudraCT Number 2009-011795-29. 2009 [Available from: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2009-011795-29/FI> accessed 1 May 2018.
5. Robinson M, Oakes TM, Raskin J, et al. Acute and long-term treatment of late-life major depressive disorder: duloxetine versus placebo. *Am J Geriatr Psychiatry* 2014;22(1):34-45. doi: 10.1016/j.jagp.2013.01.019 [published Online First: 2013/12/10]
6. López RJ, López BMA, Vargas TBE, et al. Estudio doble ciego con antidepressivo, psicoterapia breve y placebo en pacientes con depresión leve a moderada. *Salud Mental* 2004;27(5):53-61.
7. Barber JP, Barrett MS, Gallop R, et al. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *J Clin Psychiatry* 2012;73(1):66-73. doi: 10.4088/JCP.11m06831 [published Online First: 2011/12/14]
8. Dimidjian S, Hollon SD, Dobson KS, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol* 2006;74(4):658-70. doi: 10.1037/0022-006X.74.4.658 [published Online First: 2006/08/03]
9. Lecrubier Y, Bourin M, Moon CA, et al. Efficacy of venlafaxine in depressive illness in general practice. *Acta Psychiatr Scand* 1997;95(6):485-93. [published Online First: 1997/06/01]

BMJ Open

Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2018-024886.R2
Article Type:	Research
Date Submitted by the Author:	18-Oct-2018
Complete List of Authors:	Munkholm, Klaus; Nordic Cochrane Centre, Righospitalet, 7811 Paludan-Müller, Asger; Nordic Cochrane Centre, Righospitalet, 7811 Boesen, Kim; Nordic Cochrane Centre, Righospitalet, 7811
Primary Subject Heading:	Mental health
Secondary Subject Heading:	Pharmacology and therapeutics, Evidence based practice
Keywords:	Adult psychiatry < PSYCHIATRY, Depression & mood disorders < PSYCHIATRY, STATISTICS & RESEARCH METHODS

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13

Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

14

Authors

15 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a
16
17

18

Author affiliations

19 ^aNordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49

Correspondence to:

50 Klaus Munkholm, M.D., DMSc.
51 Nordic Cochrane Centre
52 Rigshospitalet, 7811
53 Blegdamsvej 9
54
55
56
57
58
59
60

1
2
3
4 200 København Ø
5 Denmark
6 Phone: +45 35457112
7 E-mail: km@cochrane.dk
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Abstract

Objectives

To investigate whether the conclusion of a recent systematic review and network meta-analysis (Cipriani et al. 2018) that antidepressants are more efficacious than placebo for adult depression was supported by the evidence.

Design

Reanalysis of a systematic review, with meta-analyses.

Data sources

522 trials (116 477 participants) as reported in the systematic review by Cipriani et al. (2018) and clinical study reports for 19 of these trials.

Analysis

We used the Cochrane Handbook's risk of bias tool and the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach to evaluate the risk of bias and the certainty of evidence, respectively. The impact of several study characteristics and publication status was estimated using pair-wise subgroup meta-analyses.

Results

Several methodological limitations in the evidence base of antidepressants were either unrecognised or underestimated in the systematic review by Cipriani et al (2018). The effect size for antidepressants versus placebo on investigator-rated depression symptom scales was higher in trials with a 'placebo run-in' study design compared to trials without a 'placebo run-in' design ($p = 0.05$). The effect size of antidepressants was higher in published trials compared to unpublished trials ($p < 0.0001$). The outcome data reported by Cipriani et al. differed from the clinical study reports in 12 (63%) of 19 trials. The certainty of the evidence for the placebo-controlled comparisons should be very low according to GRADE due to a high risk of bias, indirectness of the evidence, and publication bias. The mean difference between antidepressants and placebo on the 17-item Hamilton depression rating scale (range 0 to 52 points) was 1.97 points (95% CI: 1.74 to 2.21).

Conclusions

The evidence does not support definitive conclusions regarding the benefits of antidepressants for depression in adults. It is unclear whether antidepressants are more efficacious than placebo.

Strengths and limitations of this study

- Empirical evidence was provided showing how many biases and methodological limitations in the evidence base for antidepressants for depression affect the apparent effect size for antidepressants
- For the first time, the impact of the 'placebo run-in' study design on the apparent effect size for antidepressants compared with placebo was estimated
- We reported the effect estimate of antidepressants compared with placebo as a mean difference on the investigator-rated Hamilton depression rating scale to provide an outcome measure that can be easily interpreted by patients and clinicians
- When possible, we compared the data reported by Cipriani et al. (2018) on the outcomes of total dropouts and dropouts due to adverse events with the clinical study reports that we have previously obtained from the European Medicines Agency
- Our analyses relied on the data reported in the systematic review by Cipriani et al. (2018) and we did not perform a separate literature search and data extraction; given the methodological limitations we have identified, a reliable assessment would need to be based on clinical study reports and individual patient data

Introduction

WHO estimates that 300 million people globally suffer from depression, making depression the leading cause of disability worldwide.¹ In Denmark, 10% of all adults 25 years and older were in treatment with antidepressants in 2016.² In the US, 13% of persons 12 years and older were in treatment in 2014, making antidepressants one of the three most commonly used drug classes.¹ Prescriptions for antidepressants cost the National Health Service in the United Kingdom an estimated £267 million in 2016.³ Research that guides clinical treatment of depression therefore has a potentially important impact on millions of people and on national economies.

The recent network meta-analysis of antidepressants for depression by Cipriani et al.⁴ is the largest meta-analysis of antidepressants to date in terms of included studies and participants. It specifically aimed to inform clinical guidelines, patients, physicians and policy makers by comparing 21 antidepressants for the treatment of adults with depression. The review's primary outcomes were 'response rate' (defined as the number of participants with at least a 50% reduction on an observer-rated depression scale) and overall dropout rates. The secondary outcomes were depression symptom scores, 'remission rate' (defined as the number of participants with an observer-rated depression score below a certain threshold), and dropouts due to adverse events. Cipriani et al. found that all 21 antidepressants were more effective than placebo, whereas only two of the drugs had fewer dropouts compared with placebo. Based on these findings, they⁴ ranked the antidepressants according to 'response rate' and overall dropout rate and concluded that antidepressants were more efficacious than placebo in adults with major depressive disorder. The improvement in symptom scores they found were very similar to previous meta-analyses (Figure 1), some of which have concluded that the benefit of antidepressants is doubtful.⁵⁻⁸ The review received widespread media coverage, largely citing it as finally putting to rest any doubts regarding the efficacy of antidepressants,^{9 10} and the message of antidepressants being effective was strongly conveyed by some of the authors in the press,⁹ adding that the benefits outweigh side effects.¹⁰

There are many methodological limitations in trials of antidepressant agents,¹¹ of which many have been acknowledged for decades.¹² Research aiming to inform clinical practice on the use of antidepressants for depression must recognise these limitations. We have already addressed some of the limitations in the risk of bias assessment in the Cipriani et al. review.¹³ However, given the potential implications of Cipriani et al.'s review,⁴ we here aimed to provide a more comprehensive assessment. Specifically, we wished to investigate how the methodological limitations in the evidence base were addressed, whether the review's assessment of the risk of bias within the included trials and the evaluation of the certainty of evidence were appropriate and followed the authors' stated methods, and whether the conclusion was supported by the evidence. We furthermore aimed to provide empirical evidence on the impact of these methodological limitations by using the data reported by Cipriani et al.⁴

Methods

Data collection

1
2
3
4 We extracted the review's risk of bias assessments and descriptive data from the online supplement
5 and converted the data to Microsoft Excel format. We downloaded the online dataset⁴ and merged the
6 files for our statistical analyses.
7

8 We cross-referenced the included trials with the clinical study reports that we previously obtained from
9 the European Medicines Agency (EMA) in 2010.¹⁴ We compared the outcomes of total dropout rates
10 and dropouts due to adverse events as reported in the clinical study reports with the data reported by
11 Cipriani et al.⁴
12

13 **Statistical analyses**

14 Descriptive analyses were made in Microsoft Excel. We used the statistical software R (version 3.4.3)
15 for random effects meta-analyses based on the inverse variance method and calculated effect sizes
16 as standardised mean differences (SMD) as Hedges' *g* with corresponding 95% confidence intervals
17 (95% CI). The extent of variation among the intervention effects observed in different studies was
18 calculated as Tau² and the percentage of the variability in effect estimates that was due to
19 heterogeneity was calculated as I². For the comparisons between antidepressants and placebo on
20 rating scales, we used the Hartung-Knapp-Sidik-Jonkman approach because it results in fewer type I
21 errors than the DerSimonian and Laird approach.¹⁵ We based our analyses on the number of
22 participants from Cipriani et al.'s 'efficacy' analyses.⁴ In studies with more than one drug arm the total
23 number of participants in the placebo group was split evenly between the active comparisons and the
24 means and standard deviations were unchanged.¹⁶ We did subgroup analyses based on the use of a
25 'placebo run-in' study design, sponsorship and publication status, according to the trial characteristics
26 published by Cipriani et al.⁴
27
28
29
30

31 **Quality assessments**

32 We evaluated whether Cipriani et al.'s risk of bias assessments were in accordance with the Cochrane
33 Handbook,¹⁶ as stated by the authors.⁴ Where the approach differed we compared the risk of bias
34 assessment by Cipriani et al.⁴ with our reassessment following the Cochrane Handbook.¹⁶ The specific
35 domains (and type of bias) assessed were sequence generation (selection bias), allocation sequence
36 concealment (selection bias), blinding of participants and personnel (performance bias), blinding of
37 outcome assessment (detection bias), incomplete outcome data (attrition bias), selective outcome
38 reporting (reporting bias) and other potential sources of bias.¹⁶
39

40 We used the Grading of Recommendations Assessment, Development and Evaluation (GRADE)¹⁷
41 approach to evaluate the certainty of evidence, which, for systematic reviews, reflects the extent of
42 confidence that an estimate effect is correct. GRADE considers five domains that affect the quality of
43 the evidence: the included trials' internal risk of bias, inconsistency of the included trials' results and
44 large heterogeneity, indirectness of the evidence due to poor external validity, imprecision of the effect
45 estimate and wide confidence intervals, and publication bias.¹⁷
46
47
48
49

50 **Patient involvement**

51 No patients were involved in the development of the research question, design and implementation of
52 the study, or interpretation of the results.
53
54
55
56
57
58
59
60

Results

Risk of bias

Randomisation sequence generation and allocation sequence concealment

Cipriani et al.⁴ judged 426 (82%) and 460 (88%) of the 522 included trials to be of unclear risk of bias with respect to randomisation sequence generation and allocation concealment, respectively. The remaining trials were of low risk of bias. Trials at high or unclear risk of bias within these two domains are likely to report inflated effect estimates, especially of subjective outcomes.¹⁸ Cipriani et al. did not describe how they assessed the risk of bias in relation to the randomisation sequence generation or the allocation concealment, and we were therefore unable to evaluate if their methods followed those outlined in the Cochrane Handbook.¹⁶

Blinding of participants, personnel and outcome assessment

Cipriani et al.⁴ did not use the standard Cochrane categorisation of low, unclear or high risk of bias due to a lack of blinding.¹⁶ They categorised instead 513 (98%) studies as 'stated-not tested' in at least one of the three blinding domains, meaning that the trial had stated to be double-blind, but did not test the blinding integrity. While this implied the presence of a blinding issue, their categorisation did not affect the overall risk of bias assessment⁴ and it seemed that the 'stated-not tested' domains were counted as 'low risk of bias'. Two of the three trials categorised by Cipriani et al.⁴ to be at low risk of bias in the blinding of participants' domain had tested the blinding integrity (S1 Appendix). The blinding was likely compromised in both trials. Adverse effects of antidepressants are common and often reveal who receives active medication and who receives placebo in a randomised trial. The degree of unblinding is extensive and leads to inflated effect estimates,¹⁹ and smaller effects have been observed when the trials were better blinded by adding atropine to the placebo.²⁰ Given these issues, all placebo-controlled trials of antidepressants should arguably be categorised as at least unclear, or perhaps even at high risk of bias.

Incomplete outcome data

Cipriani et al. rated trials that used an appropriate imputation method as low risk of bias.⁴ Trials that used an 'inappropriate' imputation method were rated according to several arbitrary cut-offs: When the dropout rates were unbalanced between the arms, defined as more than a 5% difference for the head-to-head comparisons and a 10% difference for the placebo comparisons, they were rated as high risk of bias. When the dropout rates between the arms were not unbalanced but the total dropout rate was more than 20% they were rated unclear, and if the total dropout rate was less than 20% they were rated as low risk of bias. This method is not in accordance with the Cochrane Handbook, which emphasises that it is not possible to formulate a simple rule for judging a study to be at low or high risk of attrition bias in that the risk of bias depends on several factors.¹⁶ Further, the authors did not consider the reasons for dropout, although this is also recommended by the Cochrane Handbook.¹⁶

According to Cipriani et al. 121 (23%) trials were at high risk of attrition bias, but we could not replicate these results. The overall attrition rate was more than 20% in 334 (64%) trials. Using the cut-offs defined by Cipriani et al. we found that the dropout rates were unbalanced between the arms in 202

1
2
3
4 trials (39%) and according to the methods described by Cipriani et al.⁴ they should have been rated as
5 high risk of bias unless an “appropriate imputation method” was used. Cipriani et al. characterised the
6 last observation carried forward (LOCF) method as inappropriate,²¹ but they did not provide data on
7 the used imputation method in the included trials. We were therefore not able to apply Cipriani et al.’s
8 categorisations in our reassessment of the attrition bias. Most antidepressant trials use the LOCF
9 imputation method,²² which may lead to an underestimation of the variability, a falsely low p-value, and
10 an overestimation of treatment effects.²³
11
12

13 **Selective outcome reporting**

14 Cipriani et al.⁴ judged that 402 (77%) of the 522 trials were of low risk of outcome reporting bias, 100
15 (19%) of unclear risk, and 20 (4%) of high risk of bias. Their assessments were based on the reporting
16 of the review’s two primary outcomes of ‘response rates’ and overall dropout rates and a trial was only
17 rated at high risk of bias in case both outcomes were missing. This is not in accordance with the
18 Cochrane Handbook, in which a study-level judgement across all relevant outcomes is
19 recommended.¹⁶ According to our analyses the review’s three secondary outcomes of dropouts due to
20 adverse events, depression symptoms measured on depression symptom scales, and ‘remission
21 rates’ were not reported in 93 (18 %) trials, 98 (19%) trials, and 71 (14 %) trials, respectively. We
22 found that a total of 182 (35%) trials did not report at least one primary or secondary outcome and,
23 following the recommendation by the Cochrane Handbook to consider all relevant outcomes, these
24 trials should probably have been rated as high risk of bias.¹⁶ Selective outcome reporting leads to
25 overestimation of the benefits and underestimation of the harms of interventions.²⁴
26
27
28
29

30 **Other bias domain**

31 The authors omitted the ‘other bias’ domain from the risk of bias assessment although it is an
32 integrated part of Cochrane’s risk of bias tool.¹⁶ Relevant biases included in this domain were baseline
33 imbalances and design-specific risks of bias for cross-over and cluster randomised trials, which were
34 eligible according to the Cipriani et al. protocol,²¹ although the trial designs were not specified in the
35 review.⁴ Some argue that ‘vested interests’ should also be considered, since industry sponsored drug
36 studies lead to more favourable effects than other studies by mechanisms that are not explained by
37 the usual bias domains.²⁵ We explored whether industry sponsorship was associated with larger effect
38 estimates, by performing random effects meta-analyses of the placebo-controlled trials according to
39 sponsorship using the categorisation by Cipriani et al. (S1 Appendix). We found a lower effect size in
40 trials categorised as “sponsored” (SMD of 0.27 (95% CI: 0.25 to 0.30, 341 comparisons, 207 trials))
41 than in trials categorised as “unclear” (SMD of 0.39 (95% CI: 0.25 to 0.52, 12 comparisons, 10 trials))
42 and “not sponsored” (SMD of 0.41 (95% CI: 0.31 to 0.52, 37 comparisons, 36 trials) (p=0.005 for the
43 difference between the three estimates) (Table 1).
44
45
46
47
48
49

50 **Table 1.** Random effects pairwise meta-analyses of antidepressants versus placebo.

	N trials	N comparisons	ES	95% CI	Tau ²	I ²
Overall (SMD)						
Overall	253	390	0.29*	0.27-0.31	0.038	40.1%

Overall (Mean difference on the HAMD17)						
Overall	109	166	1.97**	1.74-2.21	1.896	27.6%
Publication status						
Published	196	294	0.33*	0.30-0.35	0.037	40.0%
Unpublished	57	96	0.15*	0.11-0.19	0.020	0.0%
'Placebo run-in'						
Yes	142	221	0.31*	0.28-0.34	0.043	35.0%
Unclear	79	120	0.29*	0.25-0.33	0.032	47.6%
No	30	46	0.22*	0.16-0.29	0.032	35.5%
Sponsorship						
Sponsored	207	341	0.27*	0.25-0.30	0.033	35.4%
Unclear	10	12	0.39*	0.25-0.52	0.026	33.0%
Not sponsored	36	37	0.41*	0.31-0.52	0.075	55.7%

ES: effect size; *: standardised mean difference (SMD); **: mean difference on 17-item Hamilton depression rating scale; HAMD17: 17-item Hamilton depression rating scale; 95% CI: 95% confidence interval; Tau²: estimate of overall heterogeneity; I²: inconsistency.

Summary risk of bias assessment

The authors deviated from Cochrane's overall risk of bias categorisation of low, unclear or high risk of bias,¹⁶ by introducing their own category of 'moderate' risk of bias. They classified the trials as low risk of bias if none of the domains assessed were rated as high risk of bias and three or less were rated as unclear risk; moderate if one domain was rated as high risk of bias or none were rated as high risk of bias but four or more were rated as unclear risk; and all other cases were rated as high risk of bias.⁴ This approach is similar to using scales that add up scores for multiple items to produce a total, which is discouraged in the Cochrane Handbook.¹⁶ The Handbook instead recommends an overall qualitative assessment considering the relative importance of different domains.¹⁶ The authors rated 96 (18%) of the 522 trials as low risk of bias, 380 (73%) trials as 'moderate', and 46 (9%) trials as high risk of bias. We were not able to replicate these findings and those efforts were made difficult because it was not clear how the blinding domains were rated in terms of risk of bias. Given that that the review's five outcomes were all likely affected by all of the risk of bias domains, the qualitative method suggested by the Cochrane Handbook involves classifying trials with any 'high risk of bias' domains as overall high risk of bias.¹⁶ Applying these criteria (Cochrane Handbook, Table 8.7.a)¹⁶ on Cipriani et al.'s ratings, there was one trial at low risk of bias, 383 trials (73%) at unclear risk, and 138 trials (26%) at high risk of bias. When we used our classifications for the blinding domains (i.e. all placebo-controlled trials were rated as unclear risk of bias, and for the selective outcome reporting domain) there were no (0%) trials at low risk of bias, 261 (50%) trials of unclear risk, and 261 (50%) trials of high risk of bias (S1 Appendix). If the three blinding domains were rated as high risk of bias in the placebo-controlled trials, rather than unclear risk of bias, there were no (0%) trials at low risk, 108 trials (21%) of unclear risk, and 414 trials (79%) of high risk of bias (S1 Appendix).

Publication bias

1
2
3
4 Publication bias of antidepressant trials is pervasive and distorts the evidence base.⁸ Many industry
5 funded antidepressant trials remain unpublished or are inadequately reported.⁸ Cipriani et al.⁴ included
6 436 published and 86 unpublished studies, but as many as a thousand antidepressant studies may
7 have been conducted.¹² We did a random effects meta-analysis of the placebo comparisons according
8 to publication status and found that the average effect size was lower in unpublished studies (SMD
9 0.15 (95% CI: 0.11 to 0.19, 96 comparisons, 57 trials)) than in published studies (SMD 0.33 (95% CI:
10 0.30 to 0.35, 294 comparisons, 196 trials) ($p < 0.0001$ for difference between the two estimates) (Table
11 1). Our findings are very similar to those reported by Turner et al.⁸ in 2008 of published versus
12 unpublished antidepressant trials registered by the FDA who found an SMD of 0.37 (95% CI: 0.33 to
13 0.41) for published studies and 0.15 (95% CI: 0.08 to 0.22) for unpublished studies. This indicates that
14 the reported effect sizes by Cipriani et al.⁴ are likely inflated due to publication bias. They correctly
15 downgraded their confidence in the evidence due to the risk of publication bias, but it would also have
16 been appropriate to estimate the impact of publication bias on their effect estimate.
17
18
19
20
21

22 Trial duration and long-term effects

23 Cipriani et al.⁴ extracted outcome data as close to eight weeks follow-up as possible within an interval
24 of four to 12 weeks,⁴ but they did not provide a rationale for this decision.²¹ The common clinical
25 practice is to prescribe antidepressants for much longer periods. In the Netherlands, 43% of SSRI
26 users receive treatment for 15 months or more,²⁶ while 68% of those who use antidepressants in the
27 US take them for two years or more, and 25% take them for more than 10 years.¹ Although the short
28 trial duration was acknowledged by the authors as a limitation, the lack of clinical relevance of such
29 short follow-up should have been highlighted and the confidence in the evidence should have been
30 downgraded one level in the GRADE domain of 'indirectness'. A more appropriate method would have
31 been to extract outcome data according to length of treatment and follow-up to assess any change in
32 the treatment effect over time. According to the trial characteristics reported by Cipriani et al.⁴ 12 (4%)
33 of the 304 placebo-controlled trials lasted more than 12 weeks. However, we found that only four of
34 these 12 trials contained an uninterrupted double-blind, placebo-controlled phase of more than 12
35 weeks (S2 Appendix). The two placebo-controlled trials with the longest follow-up included 81
36 participants at 36 weeks (S2 Appendix). A further consequence of a short follow-up period is an
37 underestimation of serious and non-serious adverse events.²⁷
38
39
40
41
42

43 'Placebo run-in' and inclusion of already treated patients

44 The 'placebo run-in' study design distorts the estimates of benefits and harms (Table 2A). Cipriani et
45 al. did not provide a clear definition of a 'placebo run-in',²¹ but they characterised 260 (50%) of the 522
46 included trials as having a 'placebo run-in', 182 (35%) trials as unclear, and 80 (15 %) trials as having
47 no 'placebo run-in'.⁴ We performed random effects meta-analyses of the placebo-controlled trials
48 according to the use of a 'placebo run-in' design and found that the effect sizes differed between the
49 groups with an SMD of 0.31 (95% CI: 0.28 to 0.34, 221 comparisons, 142 trials) in trials with a
50 'placebo run-in', an SMD of 0.29 (95% CI: 0.25 to 0.33, 120 comparisons, 79 trials) where the use of a
51 'placebo run-in' was unclear, and an SMD of 0.22 (95% CI: 0.16 to 0.29, 46 comparisons, 30 trials) in
52 trials without a 'placebo run-in' ($p = 0.05$ for the difference between the three estimates). In a further
53 subgroup analysis of unpublished trials without 'placebo run-in' the effect size was very small (SMD
54
55
56
57
58
59
60

0.08, 95% CI: -0.27 to 0.11, 8 comparisons, 5 trials). The use of the 'placebo run-in' design and its implications were not discussed by Cipriani et al.⁴

Table 2. 'Placebo run-in', minimal clinically significant difference, and 'response' as an outcome.

A. 'Placebo run-in' and the inclusion of already treated participants distort the benefit-harm balance

Cipriani et al.⁴ did not provide a definition of 'placebo run-in', but it usually involves that the participants, before the randomisation, receive placebo, typically for about a week after which non-adherent participants and those who responded well to the placebo (often called "placebo-responders") are excluded from the trial. Participants already in treatment with antidepressants, including the study drug, are virtually always allowed to enter the trial, and commonly all participants are tapered off ongoing antidepressant medication during the 'placebo run-in'. This study design may impact the effect estimates of placebo-controlled trials and the benefit/harm balance through several mechanisms that favour the drug over placebo:

- Participants treated with the study drug, or a similar drug, prior to inclusion and subsequently randomised to the drug will most likely tolerate it and experience fewer harms compared to a drug naïve population (**reduced harms in the drug group**).
- Participants treated with an antidepressant before the trial and subsequently randomised to placebo might experience withdrawal symptoms that can be misinterpreted as signs of worsening of the depression or as adverse events.²⁸ Withdrawal symptoms typically occur within a few days after discontinuation but there is great clinical variation²⁸ (**reduced benefits and increased harms in the placebo group**).
- Participants already treated with an antidepressant and subsequently randomised to the study drug might experience withdrawal symptoms during the placebo run-in that are alleviated by the study drug.²⁸ It could be misinterpreted as an improvement of the depression (**increased benefits in the drug group**).

B. 'Response rates' lack clinical meaning

The 'response rate' is usually defined as the number of participants in a randomised clinical trial who achieve a reduction of more than 50% of the total score on a standardised observer-rated scale for depression, such as the Hamilton depression rating scale or the Montgomery-Åsberg rating scale. 'Non-response' does not necessarily imply that the participant's condition has not improved, but simply that the improvement is rated to be less than the 50% reduction. The difference might be as little as one point on the rating scale between a 'responder' and a 'non-responder'. Thus, participants classified as 'non-responders' may actually have shown substantial improvement. The difference in 'response rates' between antidepressants and placebo does

1
2
3
4
5 therefore not indicate the difference in the number of participants who have improved, but only the
6 difference in the number of participants whose improvement exceeded the arbitrarily defined
7 threshold. In addition, by focusing on the number of participants crossing the 50% reduction
8 threshold the participants whose conditions deteriorate during the trial are ignored. Therefore, it
9 seems more clinically meaningful to look at the average effect estimate of the drug compared to
10 placebo.
11

12 13 **C. Minimal clinically relevant difference** 14

15 Cipriani et al. reported an overall effect estimate measured as a standardised mean difference of
16 0.3 between antidepressants and placebo.⁴ The British' National Institute of Health and Clinical
17 Excellence suggested in 2004 a difference of three points on the Hamilton depression rating scale,
18 or a standardised mean difference of 0.5, as a clinically significant change.⁵ However, this
19 difference was arbitrary and not based on empirical data.²⁹ Leucht et al. used clinical trial data in
20 2013 to suggest that clinicians are unable to detect reductions on the Hamilton depression rating
21 scale of three points or less.³⁰ Others have interpreted the same data and suggested that changes
22 of seven points or more on the Hamilton scale, corresponding to a standardised mean difference of
23 at least 0.875, are necessary for a clinician to detect a minimal clinical improvement.³¹ We found
24 that the mean difference between antidepressants and placebo on the 17-item Hamilton
25 depression rating scale (range 0 to 52 points), based on Cipriani et al.'s data,⁴ was 1.97 points.
26
27
28
29
30
31

32 **Dropout as a proxy for harms**

33 Overall dropout rates and dropouts due to adverse effects were assessed by Cipriani et al. as
34 measures of "acceptability" and "tolerability", respectively, whereas the antidepressants' actual harms
35 and serious and non-serious adverse events were not assessed. It can be meaningful to use total
36 dropout rates as a measure of the overall benefit/harm balance, but due to the biases introduced by
37 including participants who are already known to tolerate an antidepressant drug and the use of a
38 'placebo run-in', this outcome will likely be biased in favour of the active drug (Table 2A). Furthermore,
39 by not including a careful analysis of the serious harms, which include aggression, suicide and
40 death,³² and of specific adverse events, the review provided no basis for balancing the benefits and
41 harms, which is essential for informed consent and shared clinical decision-making and for evaluating
42 the drugs' clinical value. Adverse effects of antidepressants are common and a recent meta-analysis
43 of 131 trials of SSRIs for depression found an increased risk of serious adverse events compared to
44 placebo (OR 1.37; 95% CI: 1.08 to 1.75).⁶ This is likely an underestimate, as only 44 of the 131
45 included trials reported these data⁶ and as serious harms, including death, of antidepressants are
46 often not reported in published papers.³³
47
48
49

50 Except for two drugs none of the included antidepressants had statistically significant lower total
51 dropout rates than placebo.⁴ However, Cipriani et al. likely underestimated the antidepressants' total
52 dropout rates since they were missing in 58 (11%) of the trials and the dropout rates due to adverse
53 events were missing in 93 (18%) of the trials. A meta-analysis of dropouts in 73 trials based on clinical
54
55
56
57
58
59
60

1
2
3
4 study reports obtained from drug regulators, rather than published data, showed that 12% more
5 participants dropped out on antidepressants than on placebo.³⁴
6
7

8 We had access to the clinical study reports for 19 of the 522 trials included in Cipriani et al.'s review.
9 The outcomes of total dropout rates and dropout rates due to adverse events were fully reported in all
10 19 clinical study reports. In comparison with those data, total dropout rates or dropouts due to adverse
11 events were either not reported or incorrectly reported by Cipriani et al. in 12 (63%) of the 19 trials:
12 total dropout rates were not reported for two trials and incorrectly reported for seven trials; dropouts
13 due to adverse events were not reported for five trials and incorrectly reported for three trials (S1
14 Table).
15
16

17 18 Lack of patient relevant outcomes

19 Patient relevant outcomes such as quality of life and sick leave are rarely measured and reported in
20 psychiatric drug trials. Instead, the trials mostly rely on investigator-rated symptom scores, although
21 self-rated symptom scales also exist. In a systematic review of SSRIs for depression in adults, only six
22 of 131 trials reported quality of life data⁶ and even clinical study reports are unreliable because of
23 selective reporting of this outcome.³⁴ The inability to cope with daily activities and the drugs' side-
24 effects may be more important to patients than their depressed mood³⁵ and the exclusion of patient-
25 relevant outcomes in the protocol²¹ is a major limitation of the evidence and of Cipriani et al.'s overall
26 conclusion.⁴
27
28

29 30 Clinically irrelevant efficacy outcomes

31 The network meta-analysis' primary efficacy outcome was 'response rate' (Table 2B). It is a
32 problematic outcome because it lacks clinical relevance and it may create an illusion of clinical
33 effectiveness.³⁶ Dichotomisation of outcomes measured on rating scales leads to loss of statistical
34 power, and it increases the risk of false positive results³⁷ and spuriously inflated effect sizes.³⁶
35 Therefore, methodologists discourage the use of such dichotomised outcomes and they should
36 generally be avoided when rating scale data are available.³⁷ These issues also apply to the review's
37 secondary outcome of 'remission rates'. The choice made by Cipriani et al.⁴ to report only the relative
38 odds ratios and not the trials' absolute 'response rates' has been criticised.³⁸ However, even the
39 absolute 'response rates' are of limited clinical relevance. Cipriani et al.⁴ did not address the problems
40 related to 'response' and 'remission rates'.
41
42
43
44

45 46 Statistical versus clinical significance

47 Cipriani et al.⁴ also reported the standardised mean difference (SMD) on symptom rating scales, which
48 is more meaningful than the dichotomised outcomes.^{36 37} They reported an overall SMD for
49 antidepressants versus placebo of 0.30 (95% credible interval: 0.26 to 0.34), but the number of trials
50 and comparisons were unclear.⁵ We found a similar overall SMD for antidepressants versus placebo
51 for the direct pair-wise comparisons of 0.29 (95% CI: 0.27 to 0.31, 390 comparisons, 253 studies)
52 (Table 1). These effect estimates are statistically significant, but likely below what could be considered
53 a clinically relevant effect (Table 2C). We also calculated an overall mean difference for the trials that
54
55
56
57
58
59
60

1
2
3
4 reported endpoint or change scores on the 17-item Hamilton depression rating scale, which was the
5 most commonly used scale in the included trials (S2 Table). The mean difference between
6 antidepressants and placebo was 1.97 points (95% CI: 1.74 to 2.21, 166 comparisons, 109 trials) on
7 the 17-item Hamilton depression rating scale (range 0 to 52) (Table 1). This mean difference on the
8 Hamilton scale is likely also below what could be considered a clinically relevant effect (Table 2C).
9 Cipriani et al. did not discuss the clinical significance of their reported effect size.⁴
10
11

12 Selected, non-representative study populations

13 Antidepressant trials typically have extensive exclusion criteria that limit their external validity. These
14 include psychiatric comorbidities, alcohol abuse, long duration of illness, and 'non-response' to
15 previous antidepressant treatment.³⁹ The majority of patients in a clinical setting would not be eligible
16 to enter randomised trials due to such exclusion criteria,⁴⁰ and the evidence coming from these trials is
17 therefore of limited relevance. Furthermore, the exclusion of previous 'non-responders' and inclusion
18 of those who are expected to respond more favourably to treatment may bias the trials (Table 2A).
19 These issues were not considered by Cipriani et al.⁴ but should arguably have resulted in downgrading
20 of the confidence in the evidence in the GRADE domain of indirectness.¹⁷
21
22
23
24

25 The certainty of the evidence

26 Cipriani et al.⁴ assessed the certainty of evidence for the two main outcomes using the GRADE
27 approach adapted for network meta-analyses. They provided the GRADE results for the head-to-head
28 comparisons, but we were unable to find the results for the placebo comparisons.⁴
29 Following the issues related to the quality of the evidence, the certainty of evidence for the placebo
30 comparisons should arguably be downgraded two levels due to a 'high risk' of bias, two levels in the
31 domain of 'indirectness' due to short trial lengths, strict inclusion criteria and the use of 'placebo run-
32 in', in addition to downgrading one level due to publication bias as acknowledged by Cipriani et al.⁴
33 Downgrading due to the indirectness of the network meta-analysis' methodology should also be
34 considered.⁴¹ Taken together, the certainty of evidence should be 'very low'.¹⁷
35
36
37
38

39 Discussion

40 We have identified several important biases that were not taken into account in the systematic review
41 by Cipriani et al.⁴ We showed that the reported effect of antidepressants over placebo measured on
42 depression rating scales was small and likely inflated by several methodological limitations in the
43 trials. For the first time, we showed that the 'placebo run-in' study design appears to work towards
44 producing inflated effect sizes, in addition to publication bias and other methodological limitations.
45 Further, we showed that the outcome data reported by Cipriani et al. differed from the clinical study
46 reports and that their risk of bias assessment did not follow the methods outlined in the Cochrane
47 Handbook. Finally, we found that the certainty of evidence for antidepressants versus placebo for all
48 outcomes assessed should be very low. Taken together, the evidence does not support definitive
49 conclusions regarding the efficacy of antidepressants for depression in adults, including whether they
50 are more efficacious than placebo for depression.
51
52
53
54
55
56
57
58
59
60

1
2
3
4 Previous meta-analyses (Figure 1) have found similar improvement in symptom scores as Cipriani et
5 al.⁴ Several of these reviews have considered carefully the methodological limitations, assessed the
6 harms, and have drawn different conclusions.⁵⁻⁷ We found that Cipriani et al. did not assess the risk of
7 bias in accordance with the Cochrane Handbook as stated ⁴ and their results were non-transparently
8 presented. While the authors should be commended for sharing their data, most of the review's results
9 cannot be reproduced because basic information, such as the number of included studies, arms and
10 participants for each meta-analysis, was not reported. The network meta-analysis methodology may
11 hold some promise, but only in areas where clearly effective interventions exist and need to be
12 ranked., and the many statistical options should never overshadow an initial critical assessment of the
13 evidence and a clear presentation of the results. It seems misleading to rank the antidepressants
14 when we have very low confidence in the evidence. Interestingly, our pairwise meta-analysis of
15 improvement on symptom scores yielded very similar results to those reported by Cipriani et al. The
16 added benefit of the network meta-analysis methodology therefore seems unclear.⁴
17
18
19
20

21 We found that the evidence base consists of mainly short-term trials (12 weeks or less) with no
22 evidence for treatment beyond 36 weeks although most patients are treated for years.^{1 26} Further, the
23 apparent effect of antidepressants reported in the review by Cipriani et al.⁴ measured on investigator-
24 rated symptom scales was small and likely not clinically relevant. Observational studies also indicate
25 that the effectiveness of antidepressants in practice is very low: In the large, publicly funded,
26 Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, only 3% of the 4041
27 enrolled patients were considered "in remission" after one year.⁴² The recent finding based on clinical
28 study reports of randomised trials that more participants drop out on antidepressants than on
29 placebo,³⁴ further suggests that the benefits of antidepressants may not outweigh the harms.
30
31
32

33 Our findings showed that Cipriani et al.'s data⁴ were inaccurate, and their estimates may therefore be
34 incorrect because they relied on published data. It may be perceived as a limitation that we relied on
35 the data by Cipriani et al. and did not perform our own separate systematic literature search and data
36 extraction. Considering the multiple methodological limitations we have identified, it would be
37 necessary to analyse data based on clinical study reports and individual patient data to make a
38 reliable assessment of the benefits and harms of antidepressants because they are the most reliable
39 source of trial data.⁴³ There are also some limitations to our sponsorship subgroup analysis: While
40 industry sponsored studies have been found to report favourable efficacy results more often than non-
41 industry sponsored studies,²⁵ our analysis showed that industry sponsored trials reported a lower
42 effect estimate of antidepressants compared to placebo than non-industry sponsored trials on
43 investigator-rated depression symptom scales. However, there were important differences between
44 the two subgroups that likely contributed to the observed difference (S1 Figure). Non-industry
45 sponsored trials were smaller and older than industry sponsored trials and almost all of the non-
46 industry sponsored trials included by Cipriani et al. were published.
47
48
49
50

51 Our results highlight that the many hundreds of placebo-controlled trials of antidepressants have not
52 addressed the most important, patient-relevant questions regarding antidepressants' benefits and
53 harms. Although this has been known for years,¹² it has not led to changes in research practice.
54
55
56
57
58
59
60

1
2
3
4 Erroneous conclusions that antidepressants are efficacious for depression have the effect that they
5 may prevent people suffering from depression from seeking other solutions to alleviate their condition,
6 such as psychotherapy and dealing with psychosocial stressors, and they may stall funding and
7 research of such treatment modalities. Importantly, such conclusions may also lead to a loss of
8 interest in providing a better evidence base to determine the true clinical value of antidepressants.
9
10

11 Our review has two implications. First, the review by Cipriani et al.⁴ and its conclusion should be
12 carefully revisited. In the light of our findings the review should not inform clinical practice. Second, our
13 reanalysis has highlighted the need for a radical change in the way antidepressant trials are being
14 conducted, reported, and interpreted. We hope that doctors, patients, peers, and politicians will
15 consider the limitations of the current evidence of antidepressants for depression that we have
16 presented and collectively act accordingly. This involves informing the patients about the limitations of
17 the current evidence, thus providing a basis for a true informed consent, and working toward a better
18 evidence base for the use of antidepressants in the treatment of depression. To get reliable answers
19 about the antidepressants' benefits and harms in adults with depression we need large-scale,
20 industry-independent, and better blinded, long-term trials of drug naïve participants, with patient-
21 relevant outcomes rather than ranking scales.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure legends

Figure 1. Previous meta-analyses reporting effect sizes for antidepressants versus placebo in adults.

Data are reported as standardised mean differences with 95% confidence intervals.

NICE 2004⁵: SSRIs. Kirsch 2008⁷: 'New generation' antidepressants. Turner 2008⁸: All antidepressants. Arroll 2009⁴⁴: Antidepressants for depression in primary care. Data represent a pooled estimate of tricyclic antidepressants and SSRIs versus placebo, fixed effects model. Fournier 2010⁴⁵: All antidepressants. Data represent pooled estimate from three groups of severity (mild to moderate, severe, very severe), fixed effects model. Gibbons 2012⁴⁶: Fluoxetine and venlafaxine. Jakobsen 2017⁶: SSRIs. The effect size of mean change scores. Cipriani 2018⁴: All antidepressants.

Contributors

KM, ASP-M and KB contributed to the conception and design of the study. KM performed the meta-analyses. KM, ASP-M and KB analysed and interpreted the data, drafted and critically revised the manuscript and approved the final version to be published. All authors had full access to all the data and take responsibility for the integrity of the data and the accuracy of the data analysis.

Data sharing

All data files and the code for the statistical analyses are available from the Open Science Framework database (Access for peer review: https://osf.io/3prz9/?view_only=45a0afe9a09d4e9e9eb8a2c2d8d4bdc6. A DOI will be assigned after acceptance of the manuscript for publication).

Funding

The study was funded by the Nordic Cochrane Centre. The funder had no role in the design, data collection, analysis, interpretation of data, writing of the report, or in the decision to submit the article for publication. The research was designed, conducted, analysed, and interpreted by the authors entirely independently of the funding source.

Ethical approval

Not required.

Competing interests

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

Acknowledgments

We are grateful to Luis Carlos Saiz, Pharm D, PhD for valuable comments.

References

1. World Health Organisation. Media Centre. Depression: Fact Sheet. 2017 [Available from: <http://www.who.int/mediacentre/factsheets/fs369/en/> accessed 1 May 2018.
2. Danish Health Authority. MedicinForbrug - Indblik 2017: Laveste antal brugere af antidepressiv medicin de seneste 10 år [Medicine usage - 2017: Lowest number of antidepressant users in the last 10 years] 2017 [Available from: [https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/laegemidler/analyser-om-antidepressiva/laveste-antal-brugere-af-antidepressiv-medicin-de-seneste-10-aar-\(pdf\)-2017.pdf?la=da](https://sundhedsdatastyrelsen.dk/-/media/sds/filer/find-tal-og-analyser/laegemidler/analyser-om-antidepressiva/laveste-antal-brugere-af-antidepressiv-medicin-de-seneste-10-aar-(pdf)-2017.pdf?la=da) accessed 1 May 2018.
3. Yatham LN, Kennedy SH, Parikh SV, et al. Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 guidelines for the management of patients with bipolar disorder. *Bipolar Disord* 2018 doi: 10.1111/bdi.12609 [published Online First: 2018/03/15]
4. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
5. National Institute for Health and Care Excellence. Depression: Management of Depression in Primary and Secondary Care. Clinical Practice Guideline Number 23. London: National Institute for Health and Clinical Excellence; 2004 [Available from: <http://webarchive.nationalarchives.gov.uk/20080612001911/https://www.nice.org.uk/guidance/index.jsp?action=download&o=29617> accessed 1 May 2018.
6. Jakobsen JC, Katakam KK, Schou A, et al. Selective serotonin reuptake inhibitors versus placebo in patients with major depressive disorder. A systematic review with meta-analysis and Trial Sequential Analysis. *BMC psychiatry* 2017;17(1):58. doi: 10.1186/s12888-016-1173-2 [published Online First: 2017/02/10]
7. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5(2):e45. doi: 10.1371/journal.pmed.0050045 [published Online First: 2008/02/29]
8. Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358(3):252-60. doi: 10.1056/NEJMsa065779 [published Online First: 2008/01/18]
9. The Guardian. The drugs do work: Antidepressants are effective, study shows: The Guardian; [updated 21 February 2018. Available from: <https://www.theguardian.com/science/2018/feb/21/the-drugs-do-work-antidepressants-are-effective-study-shows> accessed 1 May 2018.
10. TIME. These antidepressants are most effective, study says Time Magazine: Time Magazine; [updated 21 February 2018. Available from: <http://time.com/5169013/antidepressants-more-effective-placebo-treating-depression/> accessed 1 May 2018.
11. Hengartner MP. Methodological flaws, conflicts of interest, and scientific fallacies: Implications for the evaluation of antidepressants' efficacy and harm. *Front Psychiatry* 2017;8:275. doi: 10.3389/fpsy.2017.00275 [published Online First: 2017/12/23]
12. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14. doi: 10.1186/1747-5341-3-14 [published Online First: 2008/05/29]
13. Boesen K, Paludan-Muller AS, Munkholm K. Network meta-analysis of antidepressants. *Lancet* 2018;392(10152):1011. doi: 10.1016/S0140-6736(18)31783-5 [published Online First: 2018/09/29]

14. Gotzsche PC, Jorgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686. doi: 10.1136/bmj.d2686 [published Online First: 2011/05/12]
15. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC medical research methodology* 2014;14:25. doi: 10.1186/1471-2288-14-25 [published Online First: 2014/02/20]
16. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Available from <http://handbook-5-1cochraneorg/> (accessed 1 May 2018): The Cochrane Collaboration, 2011.
17. Schünemann H, Brożek J, Guyatt G, et al. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. Available from guidelinedevelopment.org/handbook2013.
18. Page MJ, Higgins JP, Clayton G, et al. Empirical evidence of study design biases in randomized trials: Systematic review of meta-epidemiological studies. *PLoS One* 2016;11(7):e0159267. doi: 10.1371/journal.pone.0159267 [published Online First: 2016/07/12]
19. Perlis RH, Ostacher M, Fava M, et al. Assuring that double-blind is blind. *Am J Psychiatry* 2010;167(3):250-2. doi: 10.1176/appi.ajp.2009.09060820 [published Online First: 2010/03/03]
20. Moncrieff J, Wessely S, Hardy R. Active placebos versus antidepressants for depression. *Cochrane Database Syst Rev* 2004(1):CD003012. doi: 10.1002/14651858.CD003012.pub2 [published Online First: 2004/02/20]
21. 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(7):e010919. doi: 10.1136/bmjopen-2015-010919 [published Online First: 2016/07/13]
22. Woolley SB, Cardoni AA, Goethe JW. Last-observation-carried-forward imputation method in clinical efficacy trials: review of 352 antidepressant studies. *Pharmacotherapy* 2009;29(12):1408-16. doi: 10.1592/phco.29.12.1408 [published Online First: 2009/12/02]
23. Dziura JD, Post LA, Zhao Q, et al. Strategies for dealing with missing data in clinical trials: from design to analysis. *Yale J Biol Med* 2013;86(3):343-58. [published Online First: 2013/09/24]
24. Chan AW, Hrobjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291(20):2457-65. doi: 10.1001/jama.291.20.2457 [published Online First: 2004/05/27]
25. Lundh A, Lexchin J, Mintzes B, et al. Industry sponsorship and research outcome. *Cochrane Database Syst Rev* 2017;2:MR000033. doi: 10.1002/14651858.MR000033.pub3 [published Online First: 2017/02/17]
26. Huijbregts KM, Hoogendoorn A, Slottje P, et al. Long-term and short-term antidepressant use in general practice: Data from a large cohort in the netherlands. *Psychother Psychosom* 2017;86(6):362-69. doi: 10.1159/000480456 [published Online First: 2017/11/14]
27. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012;13:138. doi: 10.1186/1745-6215-13-138 [published Online First: 2012/08/22]
28. Fava GA, Gatti A, Belaise C, et al. Withdrawal symptoms after selective serotonin reuptake inhibitor discontinuation: A systematic review. *Psychother Psychosom* 2015;84(2):72-81. doi: 10.1159/000370338 [published Online First: 2015/02/28]
29. Turner EH, Rosenthal R. Efficacy of antidepressants. *BMJ* 2008;336(7643):516-7. doi: 10.1136/bmj.39510.531597.80 [published Online First: 2008/03/06]
30. Leucht S, Fennema H, Engel R, et al. What does the HAMD mean? *J Affect Disord* 2013;148(2-3):243-8. doi: 10.1016/j.jad.2012.12.001 [published Online First: 2013/01/30]

- 1
- 2
- 3
- 4 31. Moncrieff J, Kirsch I. Empirically derived criteria cast doubt on the clinical significance of
- 5 antidepressant-placebo differences. *Contemp Clin Trials* 2015;43:60-2. doi:
- 6 10.1016/j.cct.2015.05.005 [published Online First: 2015/05/17]
- 7
- 8 32. Sharma T, Guski LS, Freund N, et al. Suicidality and aggression during antidepressant treatment:
- 9 systematic review and meta-analyses based on clinical study reports. *BMJ* 2016;352:i65. doi:
- 10 10.1136/bmj.i65 [published Online First: 2016/01/29]
- 11
- 12 33. Hughes S, Cohen D, Jaggi R. Differences in reporting serious adverse events in industry
- 13 sponsored clinical trial registries and journal articles on antidepressant and antipsychotic
- 14 drugs: a cross-sectional study. *BMJ Open* 2014;4(7):e005535. doi: 10.1136/bmjopen-2014-
- 15 005535 [published Online First: 2014/07/11]
- 16
- 17 34. Sharma T. Effects of selective serotonin reuptake inhibitors (SSRIs) and serotonin -
- 18 norepinephrine reuptake inhibitors (SNRIs) on suicidality, violence, and quality of life (PhD
- 19 Thesis). Available from:
- 20 [http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_phd.p](http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_phd.pdf)
- 21 [df](http://nordic.cochrane.org/sites/nordic.cochrane.org/files/public/uploads/tarang_sharma_phd.pdf). University of Copenhagen, 2018.
- 22
- 23 35. Eiring O, Landmark BF, Aas E, et al. What matters to patients? A systematic review of preferences
- 24 for medication-associated outcomes in mental disorders. *BMJ Open* 2015;5(4):e007848. doi:
- 25 10.1136/bmjopen-2015-007848 [published Online First: 2015/04/10]
- 26
- 27 36. Kirsch I, Moncrieff J. Clinical trials and the response rate illusion. *Contemp Clin Trials*
- 28 2007;28(4):348-51. doi: 10.1016/j.cct.2006.10.012 [published Online First: 2006/12/22]
- 29
- 30 37. Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ*
- 31 2006;332(7549):1080. doi: 10.1136/bmj.332.7549.1080 [published Online First: 2006/05/06]
- 32
- 33 38. McCormack J, Korownyk C. Effectiveness of antidepressants. *BMJ* 2018;360:k1073. doi:
- 34 10.1136/bmj.k1073 [published Online First: 2018/03/11]
- 35
- 36 39. Posternak MA, Zimmerman M, Keitner GI, et al. A reevaluation of the exclusion criteria used in
- 37 antidepressant efficacy trials. *Am J Psychiatry* 2002;159(2):191-200. doi:
- 38 10.1176/appi.ajp.159.2.191 [published Online First: 2002/02/02]
- 39
- 40 40. Zetin M, Hoepner CT. Relevance of exclusion criteria in antidepressant clinical trials: a replication
- 41 study. *J Clin Psychopharmacol* 2007;27(3):295-301. doi: 10.1097/JCP.0b013e318058263f
- 42 [published Online First: 2007/05/16]
- 43
- 44 41. Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines: 8. Rating the quality of evidence--
- 45 indirectness. *J Clin Epidemiol* 2011;64(12):1303-10. doi: 10.1016/j.jclinepi.2011.04.014
- 46 [published Online First: 2011/08/02]
- 47
- 48 42. Pigott HE. The star*d trial: It is time to reexamine the clinical beliefs that guide the treatment of
- 49 major depression. *Can J Psychiatry* 2015;60(1):9-13. doi: 10.1177/070674371506000104
- 50 [published Online First: 2015/04/18]
- 51
- 52 43. Doshi P, Jefferson T, Del Mar C. The imperative to share clinical study reports: recommendations
- 53 from the Tamiflu experience. *PLoS Med* 2012;9(4):e1001201. doi:
- 54 10.1371/journal.pmed.1001201 [published Online First: 2012/04/17]
- 55
- 56 44. Arroll B, Elley CR, Fishman T, et al. Antidepressants versus placebo for depression in primary
- 57 care. *Cochrane Database Syst Rev* 2009(3):CD007954. doi: 10.1002/14651858.CD007954
- 58 [published Online First: 2009/07/10]
- 59
- 60 45. 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. doi: 10.1001/jama.2009.1943
- [published Online First: 2010/01/07]
46. Gibbons RD, Hur K, Brown CH, et al. Benefits from antidepressants: synthesis of 6-week patient-
- level outcomes from double-blind placebo-controlled randomized trials of fluoxetine and
- venlafaxine. *Arch Gen Psychiatry* 2012;69(6):572-9. doi:
- 10.1001/archgenpsychiatry.2011.2044 [published Online First: 2012/03/07]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

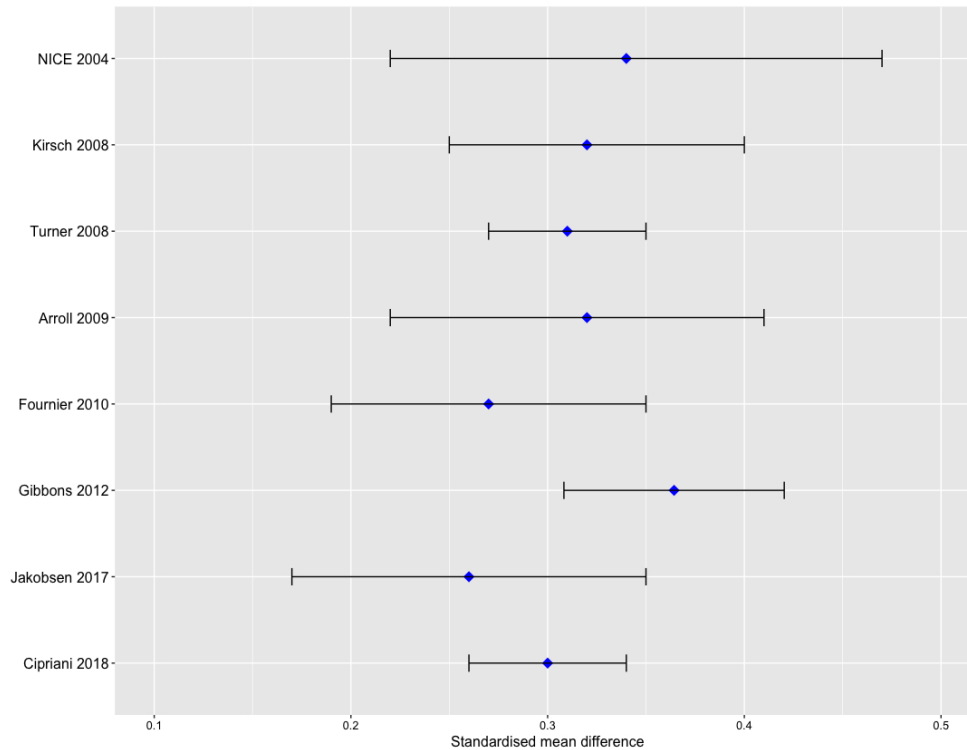


Figure 1. Previous meta-analyses reporting effect sizes for antidepressants versus placebo in adults.

Data are reported as standardised mean differences with 95% confidence intervals. NICE 2004⁵: SSRIs.

Kirsch 2008⁷: 'New generation' antidepressants. Turner 2008⁸: All antidepressants. Arroll 2009⁴⁴: Antidepressants for depression in primary care. Data represent a pooled estimate of tricyclic antidepressants and SSRIs versus placebo, fixed effects model. Fournier 2010⁴⁵: All antidepressants. Data represent pooled estimate from three groups of severity (mild to moderate, severe, very severe), fixed effects model.

Gibbons 2012⁴⁶: Fluoxetine and venlafaxine. Jakobsen 2017⁶: SSRIs. The effect size of mean change scores. Cipriani 2018⁴: All antidepressants.

389x293mm (72 x 72 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

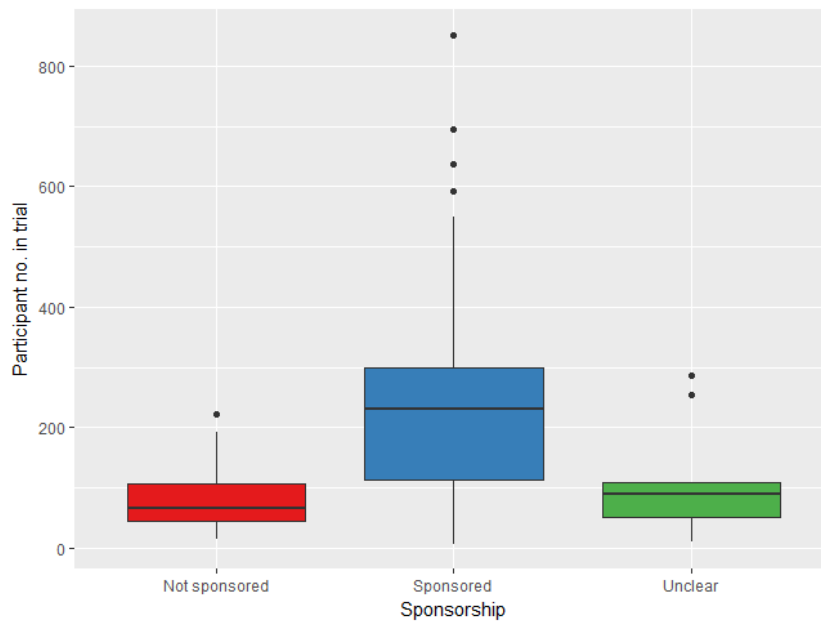
Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

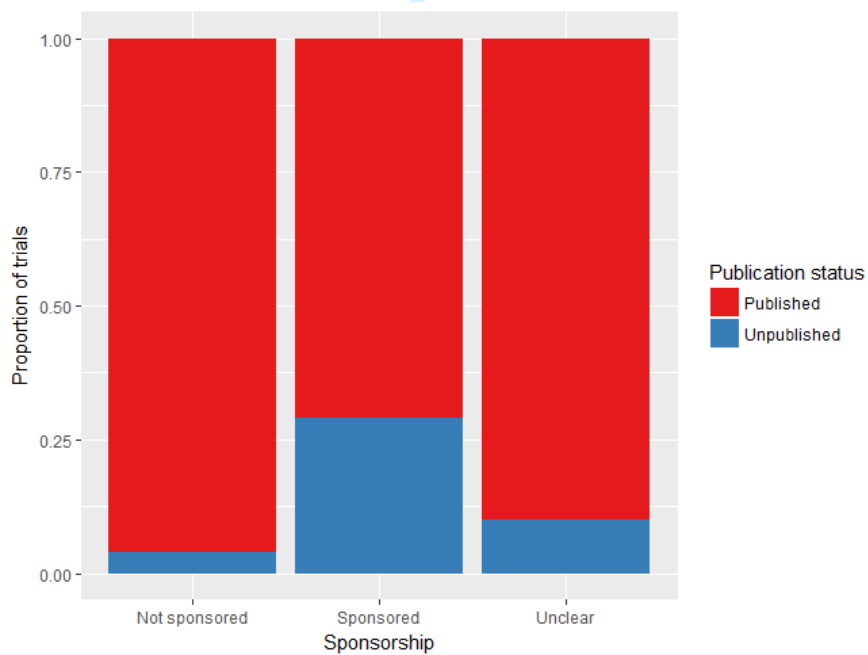
^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Figure

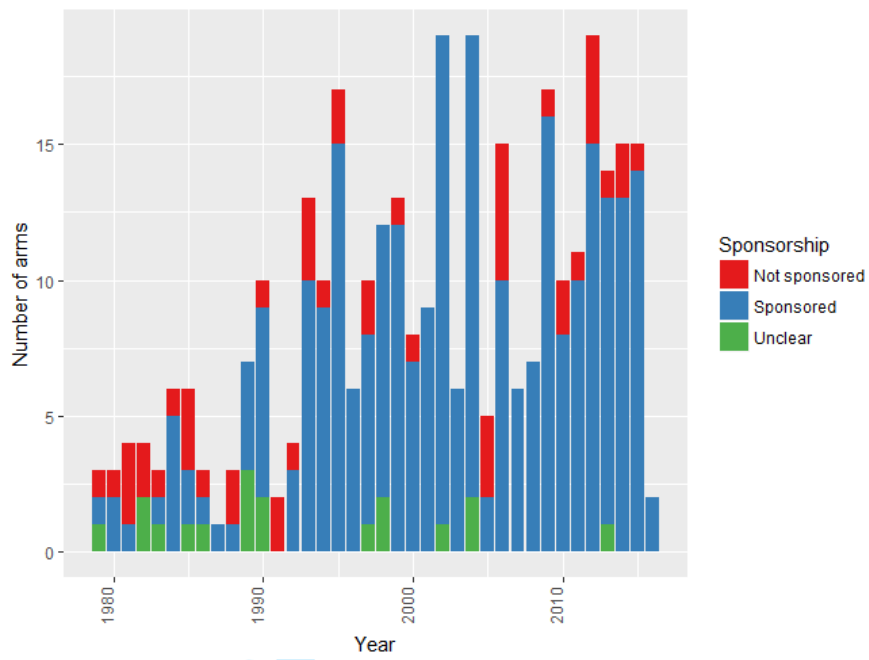
For peer review only



S1 Figure. A Size of placebo-controlled trials in the review by Cipriani et al.¹ according to sponsorship.



S1 Figure. B Sponsorship status of placebo-controlled trials in the review by Cipriani et al.¹



S1 Figure. C Sponsorship of placebo-controlled trials in the review by Cipriani et al.¹ according to publication year.

Peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Table

For peer review only

S1 Table. Comparison between the outcomes of total dropout rates and dropouts due to adverse events as reported by Cipriani et al.¹ and the data reported in the clinical study reports.

Trial ID (From Cipriani et al. dataset)	Total dropouts reported by Cipriani et al. compared with clinical study reports*	Dropouts due to adverse events reported by Cipriani et al. compared with clinical study reports**
Goldstein2002 (HMAQ - Study Group A)	Match	Match
Study F1J-MC-HMAQ - Study Group B (starting page in the pdf document: 147)	Match	Not reported
Goldstein2004a (HMAQ - Study Group A, ID#4091)	Discrepancy	Match
Goldstein2004b (HMAQ - Study Group B, ID#4091)	Match	Discrepancy
Detke2004 (HMAQ Study Group A)	Match	Match
Perahia2006 (HMAQ - Study Group B)	Match	Match
Detke2002a (HMBH - Study Group A)	Match	Match
Detke2002b (HMBH - Study Group B)	Match	Discrepancy
0600B-367	Not reported	Not reported
0600B1-384	Discrepancy	Discrepancy
VEN 600A-303 (FDA)	Discrepancy	Match
Cunningham1994 (VEN 600A-302 FDA)	Discrepancy	Match
VEN 600A-313 (FDA)	Discrepancy	Match
0600A1-372	Not reported	Not reported
Thase1997 (VEN XR 209 FDA)	Match	Match
Rudolph1998 (VEN 600A-203 (FDA)	Match	Match
SER 101 (FDA)	Match	Not reported
Fabre1995 (SER 103 FDA)	Match	Match
SER 310 (FDA)	Discrepancy	Not reported

*Our comparison considered all treatment arms in the trials. If data for all arms matched, we assigned an assessment of 'match'; if data for any arm differed but were reported, we assigned an assessment of 'discrepancy'; if data for any arm was not reported, we assigned an assessment of 'not reported'.

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]

For peer review only

1
2
3
4
5 **Considering the methodological limitations in the evidence base**
6 **of antidepressants for depression: A reanalysis of a network**
7 **meta-analysis**
8
9

10
11
12 Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

13
14 ^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark
15
16

17
18 **S2 Table**
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

S2 Table. Rating scales used in the included placebo-controlled trials (n = 304).¹

Scale	Number of trials	Number of treatment arms
HAMD17	137	198
HAMD21	92	139
HAMD24	13	22
HAMD29	1	2
HAMD31	1	2
HAMD unspecified	22	25
IDS-IVR-30	1	1
MADRS	29	50
NA	8	12

NA: not available; HAMD17: 17-item Hamilton depression rating scale; IDS-IVR-30: 30-item Inventory of Depressive Symptomatology-Self Report; MADRS: Montgomery-Åsberg Depression Rating Scale.

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]

For peer review only

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S1 Appendix

For peer review only

Risk of bias

Blinding of the participants

Cipriani et al.¹ classified three²⁻⁴ out of the 522 trials as being at low risk of bias in the “blinding of participants” domain:

Brunoni et al.² tested the blinding, by asking the participants who completed the trial to guess their allocation. They reported that 39 (75%) of 52 participants on placebo, and 29 (58%) of 50 participants on sertraline were able to correctly guess their allocation. Brunoni et al. suggested that the results were “driven by clinical improvement... rather than blinding failure”². We disagree and would categorise this trial at high risk of blinding bias of the participants.

Edwards et al.³ did not describe how and when they tested the blinding, and two of the included participants were not asked. They reported that 12 (60%) of 20 participants on placebo and 12 (63%) of 19 participants on paroxetine were able to correctly guess their treatment allocation. Edwards et al. concluded that their results “confirmed the blindness of the study”, but we would categorise the trial to be at unclear risk of bias.

Schatzberg et al.⁴ did not test the blinding, and it is unclear why this trial was rated at low risk of bias, rather than the “stated but not tested” categorisation.

Other bias domain

While Cipriani et al.¹ categorised each arm individually according to sponsorship, we considered sponsorship on the study level: trials with any sponsored arm (as categorised by Cipriani et al.¹) were categorised as “sponsored”; of the remaining trials, those with any arms categorised as “unclear” we labelled “unclear” and the remaining trials were categorised “not-sponsored”.

Summary risk of bias assessments

Criteria for assessments

To categorise the 522 trials included by Cipriani et al.¹, we followed the Cochrane Handbook’s criteria for an overall risk of bias assessment.⁵ Each domain in the risk of bias tool likely affects all five included outcomes assessed by Cipriani et al., and we therefore considered all bias domains as “key domains”, according to table 1.

Table 1. Criteria for overall risk of bias assessment.

Low risk of bias	All key domains classified as low risk of bias
Unclear risk of bias	One or more key domains classified as unclear risk of bias, and no domains classified as high risk of bias.
High risk of bias	One or more key domains classified as high risk of bias

We collapsed the three blinding domains in our Excel dataset and used the following criteria for our categorisation: All placebo-controlled trials were classified as unclear risk of bias in the main analysis and as high risk of bias in the sensitivity analysis. Trials that only contained head-to-head antidepressant arms, and no placebo arm, were rated as low risk of bias if the three blinding domains were rated as ‘low’ or ‘stated but not tested’ by Cipriani et al.¹. We rated the collapsed blinding domain as unclear risk of bias, if one or more of the blinding domains were rated as unclear by Cipriani et al. Trials with missing data for any of the five included outcomes were categorised as high risk of bias. For the remaining bias domains, we adopted the categorisations by Cipriani et al.¹. Our results are compared with Cipriani et al.’s in table 2.

Table 2. Comparison of the overall risk of bias assessments.

Cipriani et al. overall assessment		Our overall assessments			
		Cochrane categorisation	Sensitivity analysis using the Cipriani et al. categorisation	Our assessment	Sensitivity analysis of the blinding domains
Low risk	46 trials (9%)	Low	1 trial (0.2%)	0 trials (0%)	0 trials (0%)
"Moderate" risk	380 trials (73%)	Unclear	383 trials (73%)	261 trials (50%)	108 (21%)
High risk	96 trials (18%)	High	138 trials (26%)	261 trials (50%)	414 (79%)

For peer review only

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
2. Brunoni AR, Valiengo L, Baccaro A, et al. The sertraline vs. electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial. *JAMA psychiatry* 2013;70(4):383-91. doi: 10.1001/2013.jamapsychiatry.32 [published Online First: 2013/02/08]
3. Edwards JG, Goldie A. Placebo-Controlled trial of paroxetine in depressive illness. *Human Psychopharmacology: Clinical and Experimental* 1993;8(3):203-09. doi: 10.1002/hup.470080306
4. Schatzberg AF, Kremer C, Rodrigues HE, et al. Double-blind, randomized comparison of mirtazapine and paroxetine in elderly depressed patients. *Am J Geriatr Psychiatry* 2002;10(5):541-50. [published Online First: 2002/09/06]
5. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Available from <http://handbook-5-1-cochraneorg/> (accessed 1 May 2018): The Cochrane Collaboration, 2011.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Considering the methodological limitations in the evidence base of antidepressants for depression: A reanalysis of a network meta-analysis

Klaus Munkholm ^a, Asger Sand Paludan-Müller ^a, Kim Boesen ^a

^a Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

S2 Appendix

For peer review only

Trial duration and long-term effect estimates

Overall trial length

According to the trial characteristics by Cipriani et al.¹ 492 (94%) of the 522 included trials lasted between four and 12 weeks (figure 1), 28 trials (5%) lasted more than 12 weeks, and the trial duration was unclear for two trials (table 3).

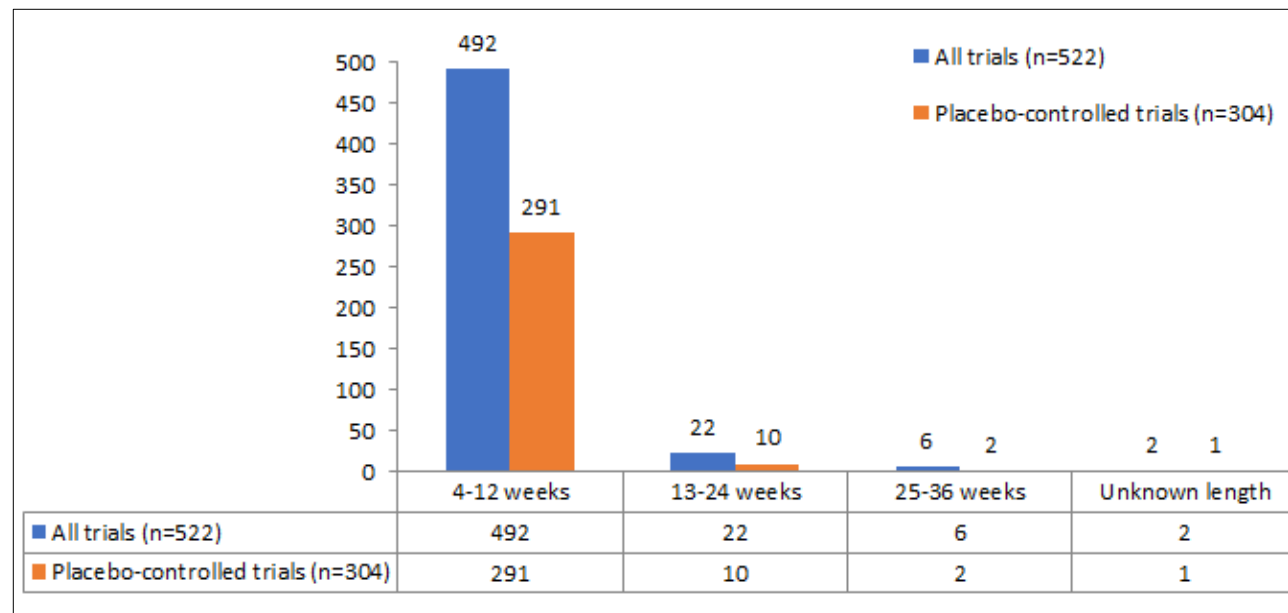


Figure 1. All 522 trials arranged into three follow-up periods according to Cipriani et al.'s trial characteristics: Four to 12 weeks (the period used by Cipriani et al.), 13 to 24 weeks, and 25 to 36 weeks¹.

Longer-term placebo-controlled trials

Of the 28 trials lasting more than 12 weeks, 12 trials had a placebo-controlled arm (table 3) and 16 were head-to-head trials. However, upon closer examination, eight of the 12 placebo-controlled trials consisted of several 'phases' and only four trials contained a continuous randomised placebo-controlled phase of more than 12 weeks (table 4). It is unclear why Cipriani et al.¹ did not list the correct trial length characteristics, since the various extension phases were clearly described in the available documents, also for the unpublished trials.

Table 1. All 522 trials arranged according to the trial length characteristics by Cipriani et al.¹

Trial length according to Cipriani et al. ¹ (N = 522)	Trials without a placebo arm (N)	Placebo-controlled trials (N)
4 weeks	9	39
5 weeks	11	4
6 weeks	92	100
7 weeks	9	0
8 weeks	59	118
9 weeks	3	2
10 weeks	4	10
12 weeks	14	18
13 weeks	1	1
16 weeks	2	2
20 weeks	0	1
24 weeks	9	6

25 weeks	2	0
26 weeks	1	0
36 weeks	0	2
Trial length unknown	2	1
Total	218	304

Table 2. Longer-term placebo-controlled trials (trial length of more than 12 weeks).

Trial	Drug arm 1	Drug arm 2	Trial length according to Cipriani et al. (weeks)	Actual trial length (weeks)	Study design and phases
Oakes 2012a ²	duloxetine	-	36	36	-
Oakes 2012b ²	duloxetine	-	36	36	-
CL3-20098-022 ³	agomelatine	fluoxetine	24	6 + 18	"Optional double-blind placebo-controlled extension" (page 24) for "responders to treatment" (page 26)
CL3-20098-023 ³	agomelatine	paroxetine	24	6 + 18	"Optional double-blind placebo-controlled extension" (page 24) for "W6 responders" (page 26)
CL3-20098-024 ³	agomelatine	fluoxetine	24	6 + 18	"Optional double-blind placebo-controlled extension" (page 24) for "responders to treatment at W6" (page 26)
CL3-20098-026 ³	agomelatine	-	24	6 + 18	"18-week extension period" of "responders at W6" (page 42)
CL3-20098-070 ⁴	agomelatine	-	24	8 + 16	16-week extension period for responders
Robinson 2014 ⁵	duloxetine	-	24	12 + 12	From week 12 to 20 placebo rescue or dose increase were available
Lopez-Rodriguez 2004 ⁶	fluoxetine	-	20	20	-
Barber 2011 ⁷	sertraline	-	16	8 + 8	Non-responders on sertraline were switched to venlafaxin ER at week 8
Dimidjian 2006 ⁸	paroxetine	-	16	8 + 8	The blind was broken at week 8 and the placebo arm was offered other treatments
Lecrubier 1997 ⁹	venlafaxine	-	13	13	-

References

1. 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 doi: 10.1016/S0140-6736(17)32802-7 [published Online First: 2018/02/27]
2. Oakes TM, Myers AL, Marangell LB, et al. Assessment of depressive symptoms and functional outcomes in patients with major depressive disorder treated with duloxetine versus placebo: primary outcomes from two trials conducted under the same protocol. *Hum Psychopharmacol* 2012;27(1):47-56. doi: 10.1002/hup.1262 [published Online First: 2012/01/14]
3. CHMP assessment report for Thymanax. EMEA/H/C/000916. 2008 [Available from: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/000916/WC500038315.pdf accessed 1 May 2018.
4. EudraCT Number 2009-011795-29. 2009 [Available from: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2009-011795-29/FI> accessed 1 May 2018.
5. Robinson M, Oakes TM, Raskin J, et al. Acute and long-term treatment of late-life major depressive disorder: duloxetine versus placebo. *Am J Geriatr Psychiatry* 2014;22(1):34-45. doi: 10.1016/j.jagp.2013.01.019 [published Online First: 2013/12/10]
6. López RJ, López BMA, Vargas TBE, et al. Estudio doble ciego con antidepressivo, psicoterapia breve y placebo en pacientes con depresión leve a moderada. *Salud Mental* 2004;27(5):53-61.
7. Barber JP, Barrett MS, Gallop R, et al. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *J Clin Psychiatry* 2012;73(1):66-73. doi: 10.4088/JCP.11m06831 [published Online First: 2011/12/14]
8. Dimidjian S, Hollon SD, Dobson KS, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol* 2006;74(4):658-70. doi: 10.1037/0022-006X.74.4.658 [published Online First: 2006/08/03]
9. Lecrubier Y, Bourin M, Moon CA, et al. Efficacy of venlafaxine in depressive illness in general practice. *Acta Psychiatr Scand* 1997;95(6):485-93. [published Online First: 1997/06/01]